Three Essays on the Impacts of Energy and Environmental Policies

by

Yiyuan Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2017

Doctoral Committee:

Assistant Professor Ying Fan, Co-Chair
Professor Ryan M. Kellogg, University of Chicago, Co-Chair
Assistant Professor Catherine Hausman
Professor Michael R. Moore

Yiyuan Zhang

yiyuanz@umich.edu

ORCID ID: 0000-0002-4574-3655

DEDICATION

To the memory of my beloved grandma.

ACKNOWLEDGMENTS

I would like to thank the Department of Economics for admitting me to the program. It has been a life-changing experience. I thank my committee members - Ying Fan, Catherine Hausman, Ryan Kellogg and Michael Moore - for their enduring support and valuable inputs. I am particularly grateful to Ryan, for many years of advising and encouragement. This dissertation would not have been possible without his guidance.

Over my years at the University of Michigan, I received tremendous help from my friends/colleagues here. I thank, among others, Paul Brehm, Maggie Brehm, Yipei Cao, Alecia Cassidy, Tong Guo, Xing Guo, Kookyoung Han, Julian Hsu, Harim Kim, Carrie Xu, Xiaoqing Zhou for collaboration and companionship. I am also indebted to my family, especially my mom, for their care, love and support for my education. I would not have reached this stage without them. Finally, my two fur-kids - Kash and Bella - have accompanied me through some of the toughest time in graduate school. They deserve a special thank you, and some extra treats as well.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

This dissertation analyzes three policies in the field of energy and environmental economics. Chapter 1 examines the impact of market design on efficiency and emissions in the wholesale electricity market. Taking advantage of Texas' transition from a decentralized bilateral trading market to a centralized auction market, I find that information aggregation has a positive effect on market efficiency that dominates any change in market power incentives. Specifically, I show that, in the nine months following the transition, high-cost generators are displaced by low-cost generators in production, leading to a total cost saving of $30.7 million relative to the counterfactual. Although the centralized market reduces generation costs, it also has an unintended effect on pollution emissions. For moderate estimates of marginal damages, I find the increase in external costs of emissions completely offsets the productive efficiency gain.

Chapter 2 examines the impact of China's fuel sulfur regulation—in particular, the introduction of CHINA III fuel standards—on air quality. Using both a time-series regression discontinuity design and a difference-in-difference method, I find that the implementation of CHINA III gasoline reduces air pollution by 4-7 percent as measured by air pollution index (API). This implies annual health benefits of 3.0 billion dollars, dwarfing the upgrading

costs of 28.2 million dollars. By contrast, the introduction of CHINA III diesel standard has only negligible impacts. Further investigation suggests that the diesel standard may not have been complied with due to the inability to pass costs through and the presence of a loophole in the diesel policy.

Chapter 3 is a joint work with Alecia Cassidy that examines the impact of the mandatory pollution liability insurance on firms' environmental compliance performance in Shenzhen, China. The mandatory pollution liability insurance is introduced as a market instrument to mitigate environmental risks. While it protects firms from liability for accidents and hence may cause moral hazard problems, firms are also incentivized to take more precautions regarding their operations and handle hazardous pollutants more properly. We seek to understand which of these effects wins out by examining the change in the number of environmental violations following the introduction of the insurance in Shenzhen. Using a novel dataset on environmental compliance performance and a triple-difference estimation strategy, we find that the number of environmental violations has decreased significantly following the passage of the insurance mandate.

CHAPTER 1

The Efficiency and Environmental Impacts of Market Organization: Evidence from the Texas Electricity Market

1.1 Introduction

How markets are organized is an important determinant of market performance. For many commodities and financial assets, markets can be organized in two basic forms: a decentralized market where transactions are conducted through private negotiations, and a centralized market where trades are intermediated by a central coordinator. For example, stocks and bonds may be traded both over-the-counter and through centralized exchanges. Given the possibility of different market forms, it is important to understand their relative merits. Starting with Wolinsky's (1990) seminal article, several theoretical studies have modeled different market forms across dimensions such as asymmetric information, search frictions, and market power (e.g., Dewatripont and Maskin (1995), Acharya and Bisin (2014), Glode and Opp (2016)). However, these studies do not provide a clear consensus regarding which market is more efficient, as it depends on the finer details of their models.

This paper adds insight into this question by providing empirical evidence on the relationship between market organization and efficiency. To do so, I focus on the US electric sector. Over the past 20 years, this sector has undergone drastic reform, as 17 states plus the District of Columbia have unbundled electric generation and retail service from transmission and distribution. In these restructured states, the wholesale electricity market takes the form of either a decentralized bilateral trading market or a centralized auction market. While a bilateral trading market relies largely on individual firms to make private transactions and dispatch decisions, a centralized market relies largely on a system operator to make scheduling and dispatch decisions based on generator bids. Among both policymakers and academics, the question of which market design supports a more efficient and competitive wholesale power market has sparked significant debate (Hogan, 1995).

To address this question, I focus on the wholesale electricity market in Texas which transitioned from a bilateral trading market to a centralized market on December 1, 2010. I examine how this market redesign affects market efficiency and social welfare. On one hand, a centralized market may improve market efficiency through information aggregation.[1] An important feature of the electricity market is the presence of network externality. It is difficult for market participants to resolve this externality in a bilateral way, due to limited information processed by each participant about others' production schedules. By contrast, in a centralized market, a system operator can utilize its central position to aggregate information from all generating units and minimize the bid-based costs. On the other hand, a centralized market may reduce efficiency if it exacerbates firms' incentive to exercise market power. In a multi-unit auction, firms have the incentive to withhold their capacity or submit bids in excess of their marginal costs to inflate the market-clearing prices. Indeed, evidence of high price-cost margins has been found in other electricity markets.[2] Therefore, whether a centralized market yields a more efficient outcome remains an open question. Moreover, changes in market organization may also affect emissions through reallocation of generating quantities among different resources. For social welfare analysis, these environmental impacts should also be taken into account along with the efficiency impact.

The core of this paper exploits hourly unit-level generation data to estimate the effect of the market redesign on generation allocation among units. The overnight change provides an opportunity to estimate the effect without contamination from changes in other aspects of the market such as generation capacity, technology and transmission capacity. These factors stayed the same within a short period preceding and following the redesign. However, demand levels and fuel prices did change even within a short window of time. I therefore rely on an econometric approach relating unit-level generation quantity to demand and fuel prices to create a credible counterfactual of generation outcomes without the market re-

[1] In studying financial markets, Acharya and Bisin (2014) propose a model in which a lack of transparency regarding trade positions leads to a counterparty risk externality. They also conclude that a centralized market improves efficiency by aggregating information about these trades.

[2] See Joskow and Kahn (2002) and Borenstein et al (2002) for evidence of market power in the California market. A growing body of research has also examined the role of transmission constraints and found that congestion adds an additional layer to the complexity of the market and thus opens up more opportunities for gaming. See Cardell et al (1997), Borenstein et al (2000), Joskow and Tirole (2000), Wolak (2015) and Ryan (2017).

design. I estimate this relationship semi-parametrically and separately for the pre-redesign and post-redesign periods, and then use the estimates from the pre-redesign period to construct the counterfactual allocation for the post-redesign period. This approach allows for considerable flexibility and avoids the needs to model the complex grid and firm behavior in detail. With the estimated changes in generation quantity for each unit and their cost and emission information, I calculate the overall cost and emission changes in this market.

The primary finding of this paper is that the centralized market improves productive efficiency. The market redesign leads to changes in generation allocation among resources of different marginal costs. As low-cost thermal generators, coal plants as a whole produce 511 more MWh per hour, which is a 3% increase in overall coal capacity utilization. The increase is significant at all levels of demand. For mid-cost combined-cycle natural gas generators, generation decreases when demand is low, which is consistent with the increase in coal. But when demand is high, there is no significant change for combined-cycle generators, while the effect on high-cost gas generators, i.e. combustion turbines and steam turbines, emerges. Specifically, generation from steam turbines decreases while generation from combustion turbines increases. Overall, my results show that high-cost generators are displaced by low-cost generators in the centralized market. Accordingly, the average hourly generation cost is estimated to be $5,062 lower than what would have been, for the nine months post redesign. This amounts to annual cost savings of $44.3 million, or a 0.5% decrease in the total generation cost. These findings suggest that the benefits from information aggregation outweigh potential market power changes associated with the move to a centralized market.

While my results indicate a productive efficiency gain from the transition to a centralized market, I also find a negative environmental impact from this transition. Specifically, I find that the increase in coal-based generation leads to an increase in carbon dioxide (CO_2) emissions by 351 tons per hour or 1.3 percent. Applying different estimates of the social cost of carbon from the EPA (2016), I find that the increase in external costs of CO_2 emissions completely offsets the private efficiency gain for moderate estimates of the social cost. The market redesign also introduces changes in sulfur dioxide (SO_2) and nitrogen oxides (NO_x) emissions. Overall, the market redesign is welfare-reducing if the external costs of emissions are taken into account.

This paper builds on and contributes to the market design literature, especially in the context of the electricity industry. While Joskow (2000) and Wilson (2002) provide an overview of the architecture of this industry, there is little empirical evidence on the relative performance of different organizational forms with the exception of Mansur and White (2012) and Cicala (2017). Both studies estimate the gains from trade due to expansion of centralized electricity markets. This paper differs from their work in several important ways. First, I focus on a context where a market transition does not involve any boundary change. This setting helps rule out the possibility that trading is impeded by administrative barriers across markets other than imperfect information related to network externality. Second, to the best of my knowledge, this is the first paper examining the environmental consequences of electricity market design. As my results suggest, these environmental impacts are critical for welfare evaluation. Third, I use an econometric approach that allows for considerable flexibility and requires no explicit assumptions on firm behavior or the grid. This is in contrast with Mansur and White (2012) which assumes away the presence of market power in their framework. Finally, my study provides a more nuanced understanding of the heterogeneous effects of market design across both generators and demand levels.

The rest of the chapter is organized as follows. Section 1.2 presents an overview of the US electricity market. Section 1.3 provides an example to illustrate how network externality, information aggregation, and market power can impact market efficiency. Section 1.4 discusses the data while Section 1.5 presents my empirical strategy. In Section 1.6, I present my findings. I provide a discussion of the results in Section 1.7 and conclude in Section 1.8.

1.2 Background

In this section, I provide a brief overview of the features of the electricity market which necessitate the adoption of independent system operators. I also discuss the market design of this industry including the role of independent system operators under different organizational forms. Given this background, I then introduce the event this study focuses on – a redesign of the Texas electricity market.

1.2.1 Basics of the Electricity Market

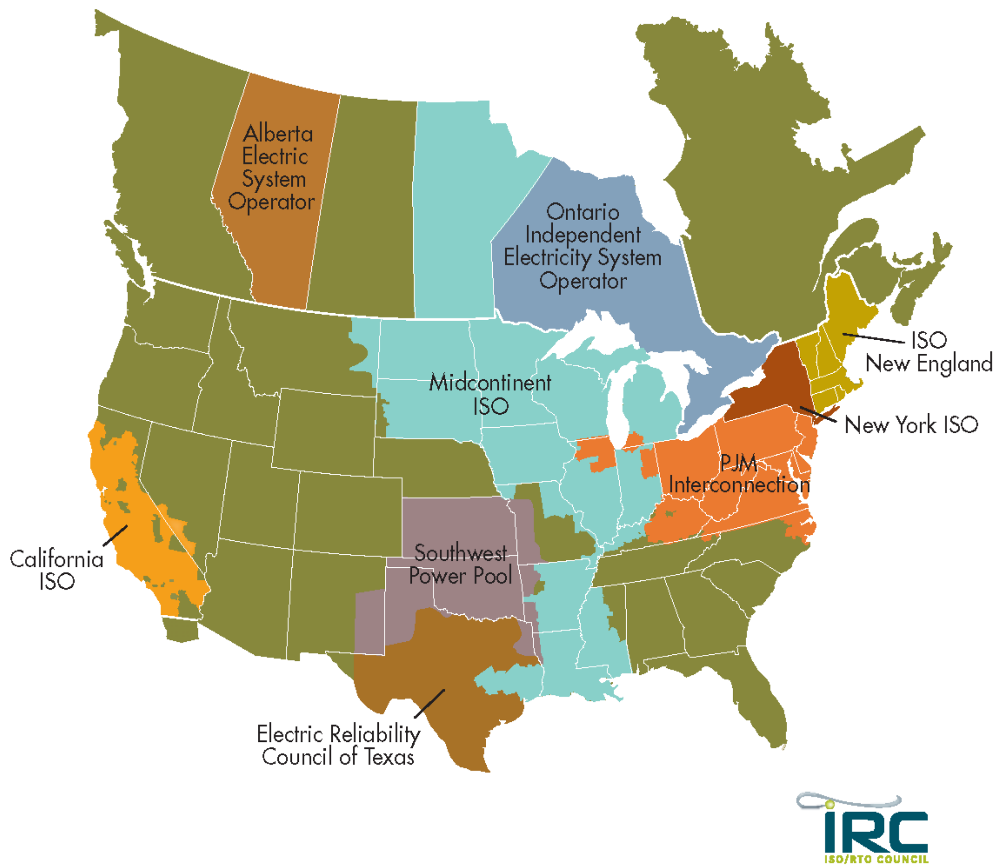
Compared to other commodities, electricity has several unique characteristics. First, the demand for electricity varies widely from hour to hour and day to day, but is almost perfectly inelastic in the short-run. That is, very few consumers are willing to or able to adjust their consumption in response to fluctuation in wholesale electricity prices. Second, electricity cannot be stored in meaningful quantities. This requires a constant real-time balance between electricity generation and consumption. Sufficient imbalances between the two can cause brownouts (a drop in electrical frequency) or blackouts (complete loss of electrical service). Third, unlike railroad networks where a supplier can designate a path for delivery, electric power flows through transmission networks according to physical laws (Kirchhoff's Laws) rather than the laws of financial contracting. Finally, the entire transmission network must meet certain physical constraints regarding frequency, voltage and capacity to ensure grid reliability.

Because of the above attributes, the proper functioning of the electricity market calls for coordination among market participants. In the US, the entire electricity market is segmented into smaller power control areas.[3] Within each power control area, an entity known as the "balancing authority" ensures both the load-generation balance and the reliability of the grid. Traditionally, vertically integrated utilities fulfill the role of balancing authorities. They own both generation and transmission assets, and hence can rely on internal scheduling and dispatch to deliver power within their exclusive service territories. While power exchanges do take place among utilities, these transactions are usually based upon mutual agreement, with each utility maintaining control over the use of its own transmission facilities.

However, since the late 1990s, several states have restructured their electric sectors and opened wholesale markets to competition. Investor-owned utilities were required to functionally unbundle their wholesale generation assets from their transmission services. To ensure open and non-discriminatory access to transmission services, FERC order 888 suggested adopting Independent System Operators (ISOs) as the balancing authorities for these restructured markets. Several ISOs emerged as a result, including the California ISO, PJM Interconnection, New York ISO, and New England ISO. These ISOs do not own any transmission assets, but exert functional control over their respective regional markets. Currently, there are 9 ISOs operating in North America, as shown in Figure 1.1.[4]

[3]A power control area (PCA) is a portion of an integrated power grid for which a single dispatcher has operational control of all electric generators. PCAs range in size from small municipal utilities such as the City of Columbia, MO, to large power pools such as PJM Interconnection. Generation and transmission facilities are physically interconnected throughout the grid, but controlled by each PCA. Since the operations of these facilities have an impact on facilities in remote control areas, the US electric industry has developed a complex set of standard operating protocols through the National Electric Reliability Council (NERC) and its eight regional reliability councils.

[4]Here, I consider ISO and RTO (Regional Transmission Organization) as synonyms.



Source: ISO/RTO Council

Figure 1.1: ISOs Operating in North America

1.2.2 Market Design of the Wholesale Electricity Market

Although the organization of each wholesale electricity market is different, the various markets can be broadly categorized into two types based on the scope of the ISO's authority and the extent of the market's centralization.

The first market design is referred to as the bilateral trading market or Min-ISO. Under the bilateral trading scheme, the role of the ISO is limited and relatively passive (Joskow, 2000). In this market, electricity buyers and sellers engage in private negotiation. The resulting bilaterally-arranged schedules are reported to the ISO. The ISO evaluates grid reliability and mitigates any energy imbalance between scheduled generation and real-time demand. This model assumes that most of the resource allocation work is done via bilateral trading, with the ISO playing only a residual balancing role. This model has been adopted by MISO (2001-2005), ERCOT (2002-2010) and CAISO (2001-2009) in the US, and NETA in the UK (2001-current).

The second market design is the centralized auction market, usually called the "electricity pool" or Max-ISO. Under this model, the ISO plays a much more active role in managing the energy market. Generation resources submit bids to supply energy to the market. The ISO then applies an optimization algorithm to the portfolio of supply offers and find the allocation with the lowest bid-based cost to achieve balance between supply and demand at every node on the network. This model has been adopted by the northeastern ISOs (NYISO, ISO-NE, PJM), MISO (2005-current), ERCOT (2010-current) and CAISO (2009-current).

1.2.3 The ERCOT Redesign

The Electric Reliability Council of Texas (ERCOT) is a nonprofit corporation certified by the Public Utility Commission of Texas (PUCT) as the independent system operator for the ERCOT region.[5] ERCOT serves 85 percent of Texas' load, 75 percent of Texas' land, and approximately 23 million customers. ERCOT is unique in that it is one of the three interconnections in North America. Limited power exchanges occur between ERCOT and neighboring regions, making it an isolated "electricity island" and thus well suited for the purposes of this study.[6]

On December 1, 2010, after years of planing, ERCOT transitioned from a bilateral trading market to a centralized auction market.[7] This transition entailed transferring most

[5]ERCOT was initially formed in 1970 to comply with NERC requirements. In 1995, the Texas Legislature amended the Public Utility Regulatory Act to deregulate the wholesale generation market, and later in 1999 passed Senate Bill 7 (SB7) to deregulate the retail electric market. Afterwards, PUCT began the process of expanding ERCOT's responsibilities to enable wholesale and retail competition and facilitate efficient use of the power grid by all market participants. On July 31, 2001, ERCOT began to operate as a single balancing authority for the entire ERCOT market, fulfilling the requirements of an ISO as specified in FERC Order 888.

[6]ERCOT is not synchronously connected to the Eastern and Western Interconnections. Power can be exchanged only via DC-ties between ERCOT and surrounding regions. There are two commercially operational DC-Ties between ERCOT and the Eastern Interconnection: North (DC_N) located near Oklaunion and East (DC_E) located near Monticello. These DC-Ties are capable of transferring a maximum power of 220 and 600 megawatts respectively. There are three additional DC-Ties connecting ERCOT and Mexico. There are no DC-Ties between ERCOT and the Western Interconnection. The overall net interchange accounts for only 0.65 percent of total net generation as of 2010.

[7]The redesign of the market was directed by PUCT in September 2003 with the goal of improving market and operating efficiencies. The initial implementation date was October 1, 2006. However, due to cost overruns and software problems, the market transition was postponed several times. The new market was

of the scheduling and dispatch responsibility from individual firms to ERCOT. Firms can rely entirely on the markets organized by ERCOT to sell and buy energy. In Appendix A, I provide more details about the scheduling and dispatch procedures under each market design.

1.3 Network Externality, Information Aggregation and Market Power

It may not be immediately clear which market scheme will produce a more efficient outcome. In this section, I examine the theoretical predictions regarding market organization and efficiency, and find that the result is indeed ambiguous. Using a simple example, I first illustrate the concept of network externality, a special form of externality in the electricity market. Then I examine how a centralized auction market can solve this externality problem and thus improve market efficiency. Finally, I show that a centralized auction market may also reduce efficiency if market power is taken into account.

1.3.1 Network Externality

As mentioned in Section 2, electricity is transmitted through an interconnected network that is subject to transmission constraints. In particular, networks can become congested. Once a network is congested, the amount of electricity that can be accommodated by the network from a particular source may depend on how much electricity is generated by other sources. This creates a special externality problem in the electricity market. Market efficiency may be impaired if market participants fail to internalize this externality.

The network externality problem can be illustrated with a simple example. Consider an equilateral triangular network with three generators. These three generators are located at the vertices, A, B and C respectively, with different marginal costs, as shown in Figure 1.2.a. All three transmission lines are identical, except that the line between A and B has a capacity limit of 100 megawatts. At point C, there is also a demand of 300 megawatts.

To meet the demand at C, the most efficient allocation is to obtain 300 megawatts from generator A, the least costly generator. The actual flow is guided by Kirchhoff's Law which states that when there are multiple paths connecting the same orientation and destination,

---

finally launched on December 1, 2010.

electrons will flow along the least resistant route.[8] Since there are two routes connecting A and C, and one of them is twice as long as the other one, the resistance of the indirect path is twice as high as the resistance of the direct path. Thus, electrons will be split in a 1:2 ratio between the indirect and direct path. Figure 1.2.a depicts the resulting electrical flows.

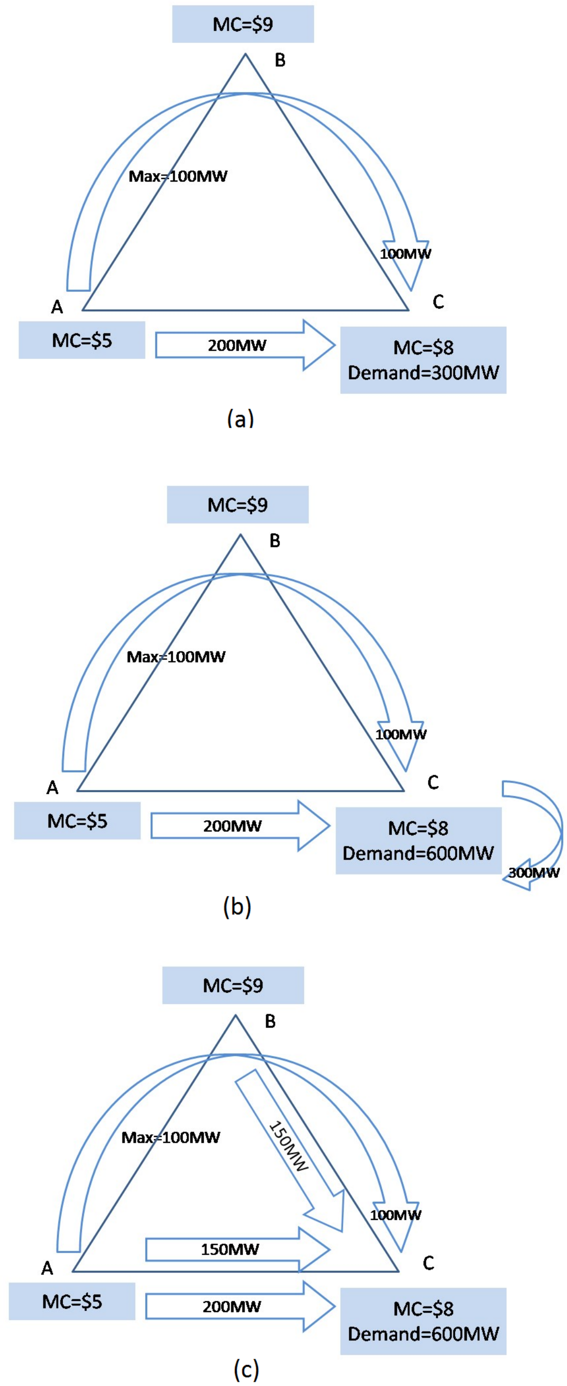
Now imagine that the demand at C is increased to 600 megawatts. At this point, the transmission line between A and B reaches its capacity limit. If generator A simply produces more electricity, some of the additional electrons will naturally flow between A and B, causing damages to the transmission line and thus reliability issues. Therefore, we must look for an alternative allocation to fulfill the increasing demand.[9] An obvious solution is to obtain additional 300 megawatts from generator C, the second least costly generator, which yields a total generation cost of \$3900. Figure 1.2.b illustrates this situation, which I refer to as the "naive allocation".

While at first glance this seems to be the best solution, the naive allocation is actually not the most efficient because it overlooks potential complementarities among generation sources in the network. Suppose we have generator B provide 150 megawatts. This at first seems to be an inefficient arrangement, since B is the most expensive generator. But generation from B alters the resistance of the indirect path from A to C. Thus, it enables greater flow from generator A through the direct path between A and C without adding extra flow through the congested path between A and B.[10] Figure 1.2.c illustrates the resulting allocation. The total generation cost is \$3600, which is lower than the generation cost under the naive allocation.

[8]The exact statement of Kirchhoff's (voltage) law is that the directed sum of the voltage around any closed network is zero. By Ohm's law, voltage is proportional to current for the same electrical circuit. Let the currents going through line $\overline{AB}$, $\overline{BC}$ and $\overline{AC}$ be I_{AB}, I_{BC} and I_{AC}, respectively. Then the combination of Kirchhoff's law and Ohm's law dictates the following relationship: $I_{AB}+I_{BC}-I_{AC}=0$. Under scenario (a), $I_{AB}=I_{BC}=\frac{1}{2}I_{AC}$.

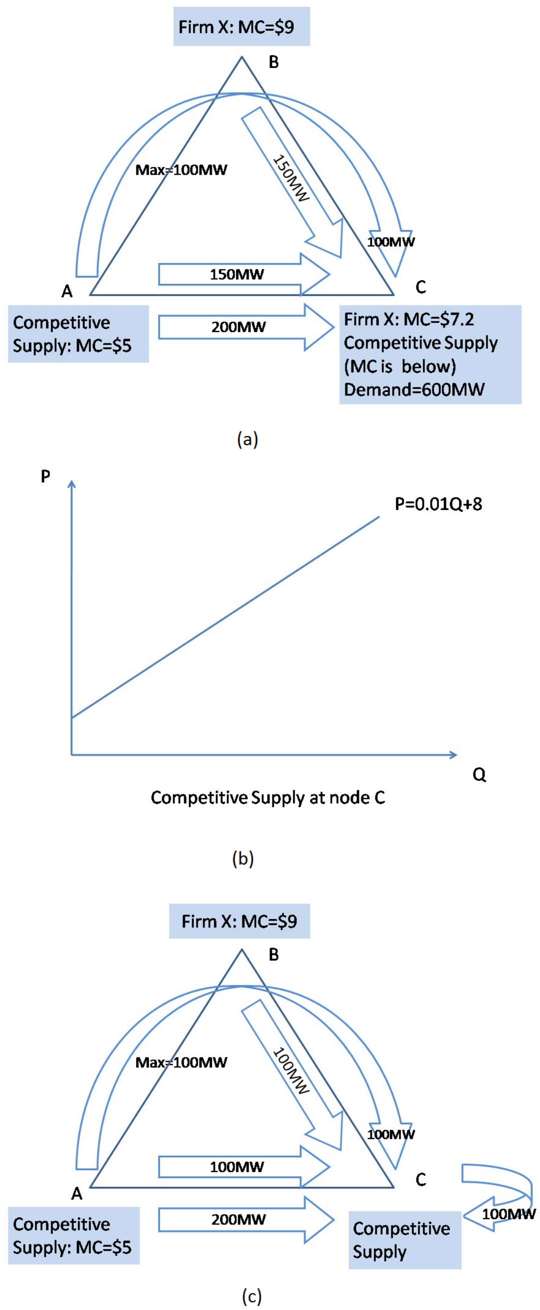
[9]One may wonder why we do not close down the link between A and B so that power can flow directly to consumers at C. Transmission lines like this are typically built for reliability reasons. For example, in case one path fails, the other provides an alternative to deliver supplies. Although transmission lines can be disconnected from the grid through a "disconnecter" or a "circuit breaker," these options are usually not intended for normal control of the circuit, but only for protection during service or safety isolation during maintenance.

[10]Recall that in footnote 8, we have $I_{AB}+I_{BC}-I_{AC}=0$. This means that we can have a higher I_{AC} by increasing I_{BC} with the same I_{AB}.



Notes: These figures provide an example illustrating the notion of network externality. Figure (a) shows the optimal allocation when demand is 300 megawatts. Figures (b) and (c) show the "naive allocation" and the optimal allocation respectively when demand is 600 megawatts. See the text for details.

Figure 1.2: An Example Illustrating Congestion Externalities



Notes: These figures provide an example illustrating the effect of market power. Figure (a) shows the market structure. Firm X has two plants located at nodes B and C. Competitive suppliers are located at nodes A and C. The supply curve of the competitive fringe at node C is given in Figure (b). The power flows drawn in Figure (a) indicate the efficient allocation when Firm X acts competitively and marginal costs are used to minimize the generation cost. Figure (c) presents the outcome when Firm X exercises its unilateral market power. See the text for details.

Figure 1.3: An Example Illustrating Market Power

1.3.2 Imperfect Information and Information Aggregation

The above example shows that finding the most efficient allocation requires knowing the marginal costs of the generators and the structure of the network, as well as being able to calculate electric flows over every segment of the transmission network.

Under a bilateral trading market, market participants possess imperfect information that prevents them from identifying network externalities and achieving efficient allocations. Although they may have a good idea of different generators' marginal costs given the similarity of the technology and the abundance of public data, they may not know others' scheduled generation since these are privately negotiated. Moreover, the actual transmission network has hundreds of lines with thousands of nodes. This complexity adds to the difficulty of identifying externalities. While it seems that market participants may learn gradually through repeated interactions, it is an illusion created by the simplicity of the network in the example. Identifying externalities in a complex network requires detailed modeling of the grid and considerable computing power. As a result, the externality problem cannot be resolved in a Coasian fashion under a bilateral trading market.[11]

By contrast, under a centralized auction market, the ISO aggregates information from generators and takes full charge of scheduling. The optimization algorithm directly takes into account the physical properties of the actual transmission lines and solves for the optimal generation allocation that minimizes bid-based generation costs over the entire network. Since network externalities are directly accounted for in the optimization procedure, the centralized market is superior at resolving the problem. Mansur and White (2012) share the same view on the source of efficiency gain.

1.3.3 Market Power

In the previous example, I assume that marginal costs are observed and used to determine economic dispatch. Since the ability of the ISO to optimize allocations depends on the

[11] Although ERCOT can re-dispatch generators in the balancing market, the adjustments do not fully correct inefficiency in the bilateral schedules, for several reasons. First, the balancing market only takes care of the imbalance between scheduled generation and demand, which is about only 5% of the overall generation, while the majority are scheduled by market participants. Second, if a bilateral schedule is feasible and yet inefficient, such as the "naive allocation" in the example, ERCOT will not change it in the balancing market. Finally, when ERCOT does change the schedules to resolve power imbalance or congestion, it adjusts generation in a piecemeal fashion, as explained in Appendix A. Using a zonal structure rather than considering the grid in its entirety, ERCOT cannot find the most efficient allocation in its adjustment.

information submitted by suppliers, inefficiency can result when firms deliberately withhold their generation capacity or submit bids substantially in excess of their marginal costs. Such behavior changes the dispatch order on the supple curve and causes high-cost generators to be used while low-cost ones stay idle.[12] Evidence of such manipulation has been found in the UK and California markets by Wolak and Patrick (2001) and Joskow and Kahn (2002), respectively.

To see how the exercise of market power reduces efficiency, consider an extended version of the previous example. As shown in Figure 1.3.a, there is a mass of competitive suppliers located at node A with a constant marginal cost of \$5. A competitive fringe supplier sits at node C with the marginal cost curve indicated in Figure 1.3.b. In addition, Firm X owns two generating units, one at node B and another at node C, with marginal costs of \$9 and \$7.2, respectively. In this example, the competitive suppliers take the strategies of Firm X as given and act as price takers.

Under this setup, the most efficient allocation remains unchanged: we should procure 450 megawatts from suppliers at A and 150 megawatts from Firm X at B. This will be the resulting allocation in a centralized auction market when marginal costs are submitted as bids, or equivalently firm X is acting competitively. The equilibrium price at point B is \$9 and the profit of Firm X is \$0.

However, suppose Firm X wishes to take advantage of its unique position in the congested network to increase its profit. If it withholds one megawatt at B, this means two additional megawatts must be generated at C, increasing prices at both B and C. In this case, Firm X has to make a classic tradeoff between profiting from a higher quantity versus a higher price. Assume Firm X uses the quantities it supplies as leverage. At the optimum, Firm X should supply 100 megawatts at point B, and 0 megawatts at point C, leaving the competitive fringe firm to meet the remaining demand, as depicted in Figure 1.3.c.[13]. The equilibrium prices at point B and C are \$13 and \$9, respectively, and Firm X's profit is \$400.[14] Relative to the competitive benchmark, Firm X is able to increase its profit by restricting its output, but the resulting allocation is no longer efficient. The total generation

[12]Note that the exploitation of market power does not necessarily indicate efficiency loss. If all firms simply bid twice their marginal costs, they will retain their order in the supply curve and hence incur no efficiency loss, despite the oligopoly rents they will enjoy. Under this scenario, the exercise of market power only causes a transfer of surplus from consumers to suppliers.

[13]See Appendix B for details.

[14]The outcome will be the same if I assume Firm X competes by bidding into the pool. The optimal strategy then is to bid \$13 and \$9 for units at B and C, respectively.

cost becomes \$3750, which is higher than not only the cost under the efficient allocation, but also the cost under the naive allocation.[15]

This example shows that the exercise of market power can significantly affect generation allocation. Therefore, whether a centralized market improves efficiency is an empirical question, depending on changes in both the management of network externality and the exercise of market power.[16] In the next sections, I will take advantage of ERCOT's market redesign to empirically examine the efficiency effect.

1.4 Data

For this study, I compiled a detailed and comprehensive dataset from a variety of sources. Most of the data are publicly-available. The sample period runs from June 1, 2010 to August 31, 2011, covering 6 months before the redesign and 9 months after the redesign.[17]

1.4.1 Generation Data

The primary data I use to determine electricity generation are from ERCOT. For each generating unit under ERCOT's control, I observe the net electrical output in 15-minute intervals.[18] I aggregate net generation at the hourly level to be consistent with the other data I have. Note that my sample is missing several days of data immediately following the market redesign, due to glitches experienced by ERCOT in that period, but is otherwise complete. Overall, there are 429 units, at 218 power plants, supplying electricity to the grid

[15]Recall that the naive allocation is the allocation when marginal costs are known but complementarity among generation sources is not taken into account. In this case, the naive allocation will procure 300 megawatts from A and 300 megawatts from Firm X at point C. The resulting generation cost for the naive allocation is \$3660.

[16]It would be interesting to directly compare market power under the two markets. Although data exist for bids submitted in centralized auction markets, price and quantity data on bilateral contracts are rarely available due to the confidential nature of these transactions. As a result, the existing literature remains silent on the market power issue in the bilateral trading setting.

[17]I exclude dates between February 2, 2011 to February 5, 2011. During this period, a strong arctic front approached Texas and resulted in the lowest temperature in 20 years. According to EIA's 2009 Residential Energy Consumption Survey, about half of Texas residents use electricity for heating. The extreme weather conditions drove up the demand for electricity. At the same time, the extreme cold also affected generation performance. More than 8,000 MW of generation unexpectedly dropped off line (40% are coal generators). The combination of these factors led to rotating outages on the ERCOT grid.

[18]A generating unit is a single turbine along with a boiler and a smokestack. A power plant usually consists of several, independently operating generating units. For combined cycle natural gas generators, however, the output decision is made jointly for both the combustion turbine and the steam turbine. Therefore, I treat them as one single unit.

managed by ERCOT.[19]

To determine ERCOT's generation portfolio, I supplement ERCOT's generation data with EIA's Annual Electric Generator Report (EIA-860 form) and Power Plant Operation Report (EIA-923 form). Table 1.1 describes the share of ERCOT's annual generation quantity and capacity by fuel type. Electricity generation in ERCOT comes almost entirely from coal, natural gas, nuclear and wind, with the rest comprising only 1% of the total generation.

Fuel Type	Share of Capacity(%)		Share of Generation(%)	
	2010	2011	2010	2011
Coal	19.95	20.34	35.15	35.61
Hydroelectric	0.58	0.55	0.28	0.15
Natural Gas	64.36	63.55	44.59	44.60
Combined Cycle	37.99	39.17	38.00	37.97
Combustion Turbine	7.13	7.06	3.52	3.72
Steam Turbine	19.24	17.32	3.07	2.91
Nuclear	5.17	5.27	12.08	10.96
Wind	9.45	9.79	6.90	7.71
Others[20]	0.49	0.50	1.00	0.97

Notes: This table reports the share of capacity and the share of generation quantity for different resource types in 2010 and 2011. Data come from EIA-860 forms and EIA-923 forms.

Table 1.1: Generation Composition in ERCOT: 2010-2011

1.4.2 Generator Characteristics

I obtain plant- and generator-level characteristics from EIA-860 forms, EPA's Continuous Emissions Monitoring System (CEMS) and eGrid.[21] For each generating unit, I observe its ownership, nameplate capacity, fuel type, technology, sector, commercial operating date,

[19]Not all the generating units in the ERCOT territory are subject to ERCOT dispatch. There are firms which provide electricity only on private networks. Nor are all the generating units dispatched by ERCOT located in Texas. In particular, the Kiamichi Energy Facility is located in Oklahoma.

For subsequent analysis, I also drop generators whose cumulative generation is less than 15 megawatt hours during the sample period. These units are not economically important.

[20]Others include biomass, petroleum coke, distillate fuel oil, solar and electricity storage.

[21]All fossil-fuel generating units with at least 25 megawatts of generating capacity have to report their hourly gross generation, heat inputs, and CO_2, SO_2, and NO_x emissions to the EPA. In the sample, 218 out of 300 thermal generators are covered by CEMS.

operating status and location, among other information. For generators with information available in CEMS, I also observe their hourly CO_2, SO_2 and NO_x emission quantities and heat inputs.

For thermal generators, I use these data to measure the average CO_2, SO_2 and NO_x emission rates and heat rates. Heat rate is the ratio of thermal energy input against electricity output. It is stable within the operating range of a generator, but can be higher during startups.[22]Heat rate reflects power plant's efficiency: the lower the heat rate is, the more efficient a generator is. For generators covered in CEMS, I estimate the average heat rate for each unit by using the slope of regressions of heat inputs (in MMBtus) on net generation (in MWhs). For generators not covered in CEMS, I use their plant-level nominal heat rates obtained from the EPA's eGrid database. To calculate average emission rates, I divide total emission quantities (in short tons or pounds) by total net generation (in MWhs).

	Coal	Natural Gas		
		Combined Cycle	Combustion Turbine	Steam Turbine
Nameplate Capacity(MW)	638.61	414.79	72.53	254.10
	(192.96)	(277.91)	(44.13)	(215.74)
Years in Operation	26.00	16.15	19.50	43.64
	(11.38)	(11.11)	(12.63)	(8.08)
Heat Rate(MMBtu/MWh)	9.55	7.38	11.90	10.99
	(0.62)	(1.35)	(4.26)	(2.82)
CO_2(Ton/MWh)	1.16	0.51	0.72	0.80
	(0.20)	(0.13)	(0.29)	(0.33)
SO_2(Pound/MWh)	5.80	0.01	0.03	0.02
	(3.86)	(0.02)	(0.05)	(0.03)
NO_x(Pound/MWh)	1.36	0.34	2.24	1.76
	(0.69)	(0.26)	(2.61)	(1.06)
Median Ramping Time[23]	Over 12H	12H	1H	12H

Notes: This table compares the characteristics of different types of thermal generators. The data for nameplate capacity, years in operation, and ramping time come from EIA-860 forms. The heat rates and emission rates are calculated by the author as described in the text. Standard deviations are reported in the parentheses.

Table 1.2: Summary Statistics of Thermal Generators' Characteristics

Table 1.2 presents the summary statistics for the thermal generators in my sample. In general, coal generators tend to be larger, more polluting and slower in ramping than natu-

[22]More than half of the startup costs are fuel costs incurred in warming up the generator. The startup cost varies by technology and unit size.

[23]Ramping time is defined as the minimum amount of time required to bring a generator from cold shutdown to full load, and is coded into four categories: 10 minutes, 1 hour, 12 hours and over 12 hours. I obtain this information from the 2013 EIA-860 form, since it is available only after 2013.

ral gas generators. Natural gas generators use several different technologies. A combustion turbine, a.k.a. gas turbine, uses high-pressure gas generated from fuel-burning to drive the turbine. A steam turbine works similarly except it uses water instead of air to drive the turbine. Most steam turbine generators currently in use were built in the 1980s and 1990s. The technology has not seen much improvement since then. By contrast, combustion turbine technology has become more efficient over time. As a result, there is a wide variation in the heat rates of combustion turbines. The heat rate of the most recent turbine is only one fourth of that of the oldest one. Besides, combustion turbines can also respond quickly to changing demand. Therefore, they are often used as peaking plants. The relatively new combined-cycle technology combines these two thermodynamic cycles together to improve the efficiency of energy conversion. As a result, they have the smallest average heat rate among natural gas generators.

1.4.3 Electricity Demand

The hourly demand data at eight weather zones are obtained from ERCOT.[24] The data is derived by aggregating meter data and includes both transmission and distribution losses. I use demand at the weather zone level to capture any spatial distribution in demand that may impact generation outcomes. In addition, I select a time period that includes both winter and summer months under each market design to capture any seasonal or diurnal patterns in demand.

1.4.4 Cost Data

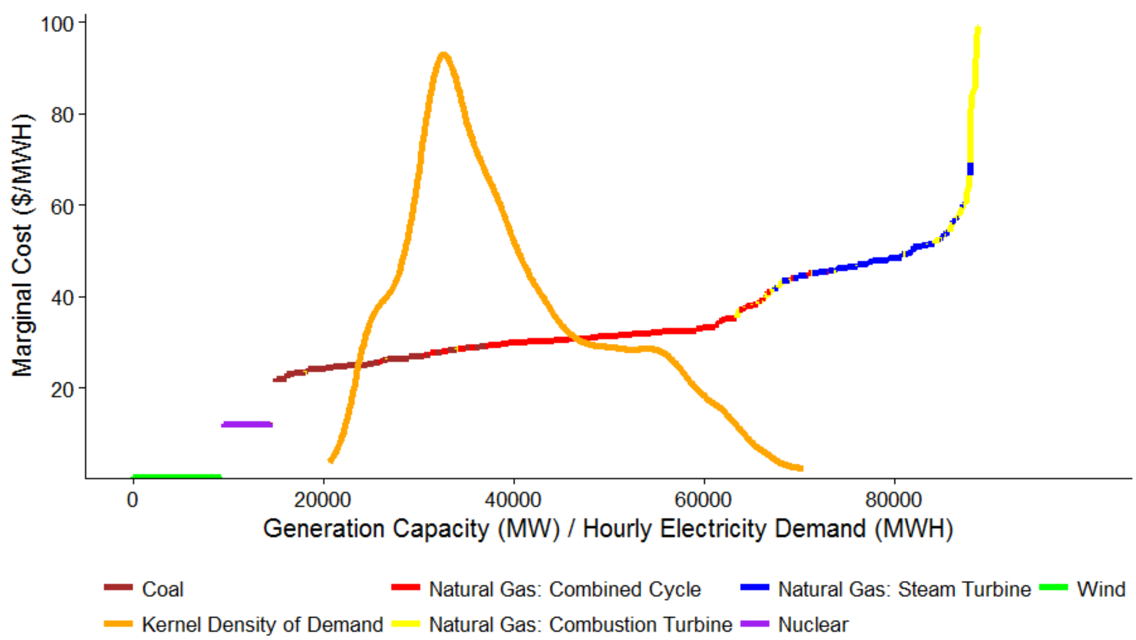
The cost structure in the electricity industry is relatively straightforward and well understood. For wind generators, the marginal cost for producing 1 more MW is essentially zero. For nuclear power plants, the marginal cost can be estimated by adding up the fuel cost and the variable operating and maintenance cost. I use the EIA (2011)'s fuel cost estimate of $7.01 and ERCOT (2012)'s VOM estimate of $5.02 to obtain a marginal cost of $12.03 for nuclear units. To estimate the marginal cost for thermal generators, I take the standard approach commonly used in the economic literature (Wolfram (1999), Borenstein et al (2002), Mansur (2008)). This methodology is based on the following elements: (1) the heat rate of each generator, (2) fuel prices, (3) variable operating and maintenance costs

[24] A weather zone is a geographic region designated by ERCOT in which climatological characteristics are similar for all areas within such a region. There are eight weather zones: coast, west, far west, east, north, north central, southern and south central.

(VOM), (4) the emission rates of each generator, and (5) emission allowance prices. Appendix C provides more details on the data sources of elements (2), (3) and (5). For each thermal generator *i* at time *t*, its marginal cost is calculated using the following formula:

$$MC_{it} = \text{Heat Rate}_i \times \text{Fuel Price}_{it} + Pno_{xt} \times NOx_i + Pso_{2t} \times SO_{2i} + VOM_i$$

Figure 1.4 plots the marginal cost curve using generators' average marginal costs and the observed maximum hourly net generation as a measure of their capacity during the sample period. Note that marginal costs differ by fuel and technology types. Specifically, wind generation and nuclear power generation are cheaper than thermal generation. Among thermal generators, coal generation is in general cheaper than natural gas generation, and combined-cycle generation is cheaper than combustion or steam generation. I also overlay on the same graph the distribution of hourly electricity demand. We can see that the marginal unit is a coal or a combined-cycle natural gas generator for most realizations of demand, while during peak hours it is a combustion or a steam turbine.



Notes: This figure plots the marginal cost curve using generators' average marginal costs and observed maximum hourly net generation as their capacity for the sample period (June 1, 2010 - August 31, 2011). It also shows the kernel density of the total hourly electricity demand for the same period. See the text for details.

Figure 1.4: Marginal Cost Curve and Distribution of Demand

1.5 Empirical Strategy

In this section, I introduce the econometric approach I use to quantify the changes in generation allocation due to the market redesign. I begin by discussing how market efficiency is measured in electricity generation.

Figure 1.5 illustrates a hypothetical efficiency change. Following the literature, I treat electricity demand as perfectly inelastic in the short-run. Demand swings are driven entirely by exogenous forces, such as temperature or human activity. Given the need to balance generation with demand at every second, there is no inefficiency from quantity distortion under either market design. Thus, any change in market efficiency will be reflected as a change in the generation cost needed to serve the same demand. The MC curve in the graph represents the theoretical efficient supply based on installed generator capacity and each generator's marginal cost. However, the actual supply will deviate from this curve. A generator operates out of the "merit order" if it is called on to help meet n megawatts of demand even though it is not the n cheapest megawatts in terms of installed capacity. Out-of-merit cost is measured as the additional production cost relative to the cost of dispatching the cheapest units. In Figure 1.5, the grey area represents out-of-merit costs under the bilateral trading market. An electricity market may incur out-of-merit costs for a number of reasons. First, transmission constraints may make it infeasible to utilize only the least-costly units.[25] Second, market participants may fail to identify and resolve network externalities. Third, they may also influence the market by exercising market power, as explained in Section 1.3. In my study, I do not focus on out-of-merit costs per se, as does Borenstein et al (2002). Instead, I focus on the difference in out-of-merit costs between the two different designs. In Figure 1.5, the centralized market brings the supply closer to the MC curve, reducing out-of-merit costs. Correspondingly, the slashed-pattern area measures the efficiency gain in this hypothetical example.

In order to measure the changes in generation costs, I need to create a credible counterfactual of what the outcome would be had ERCOT not redesigned its market. Creating this counterfactual poses several challenges. First, a market simulation with an engineering model would require modeling the electrical grid in detail and making strong assumptions

[25] Other reasons include generator outages and dynamic constraints. Plants periodically go off-line for maintenance and occasionally experience forced shutdowns, causing more expensive units to fill the gap. In addition, the start-up time and ramping costs play a role in determining the most economical dispatch, as shown by Mansur (2008) and Reguant (2014). Hence, the mere presence of out-of-merit costs does not necessarily indicate efficiency loss.

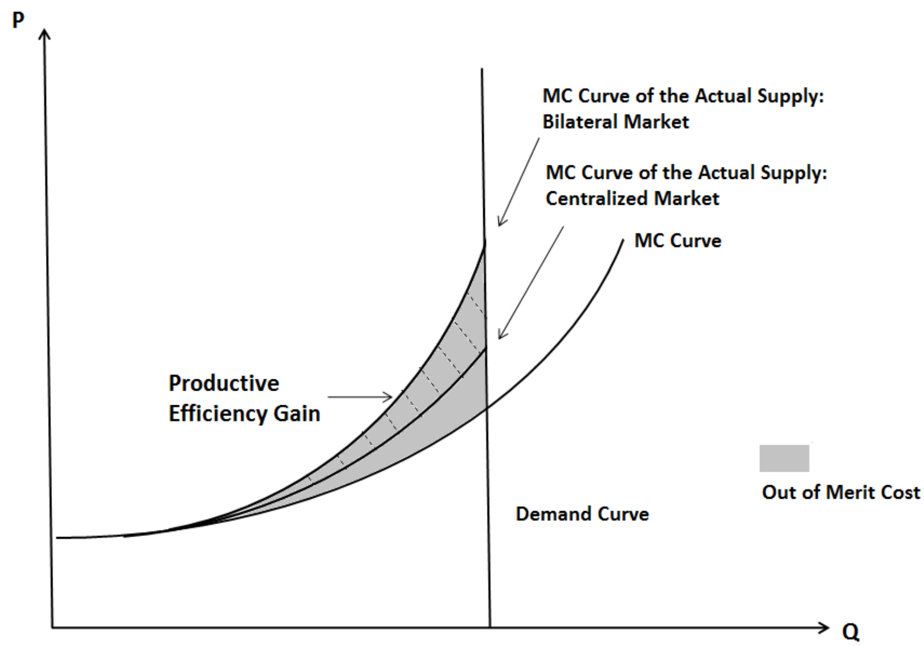
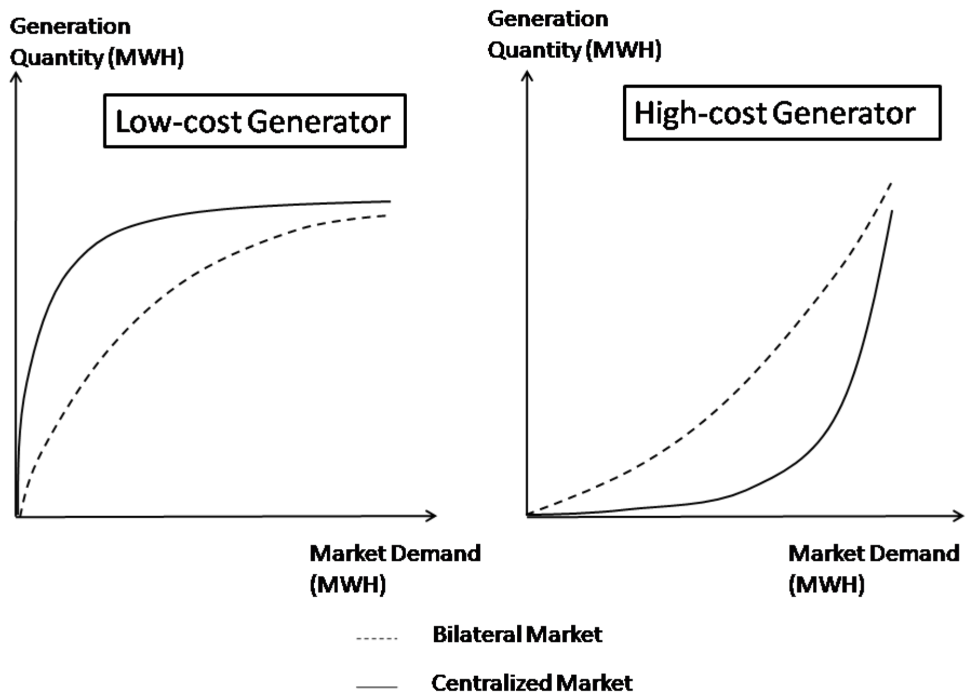


Figure 1.5: An Illustration of A Hypothetical Efficiency Gain Under the Centralized Market

regarding firms' information sets and strategic behavior. This is a difficult task given the complex nature of the transmission network as well as the competitive dynamics. Second, doing a pre/post comparison of generation costs might seem plausible as there are few changes in the market during the relatively short period preceding and following the redesign. Indeed, in Appendix D, I show that the ranges of demand levels and fuel prices are comparable pre- and post-redesign, and there is no major change in market capacity during the sample period. However, demand levels and fuel prices fluctuate even within a very short window of time. Hence, the distribution may vary over time. In particular, during my sample period, the average demand post-redesign is lower than the average demand pre-redesign. Without taking these changes into account, I may misattribute a cost reduction due to demand shifts as an indication of an effect of the market redesign.

I therefore rely on an econometric approach that estimates a flexible function of generation quantity on market demand and fuel prices for each generating unit separately before and after the market redesign. I then use these estimates to construct counterfactual generation quantities which form the basis for the calculation of changes in overall generation costs and emissions. The idea is that if the centralized market results in an improvement in market efficiency, we should in general expect an increase in electricity generation from lower-cost generators and a decrease in electricity generation from higher-cost generators, conditional on the same market demand and fuel prices. Figure 1.6 illustrates this situation. Hence, we can measure the change in generation quantity for each generator by first estimating these generation curves. This approach is similar to the one used by Davis and Hausman (2016).



Notes: Assuming constant fuel prices, this figure provides an example of changes in generation curves for generators of low versus high costs if the centralized market improves efficiency.

Figure 1.6: An Illustration of Changes in Generation Curves for Generators of Different Costs

I treat wind power and nuclear power as non-dispatchable units unaffected by the market redesign. For wind generators, generation quantity is largely determined by the availability of the wind, but it may be curtailed by ERCOT when transmission is congested. Although the market redesign can potentially improve the integration of wind power and thus reduce the incidence of curtailments, I do not find evidence supporting such a claim.[26] For nuclear power generators, given its low marginal costs and limited capacity to follow

[26] See Appendix E for more details.

load, they almost always run at full capacity unless having an outage. Based on the above reasons, the rest of my analysis will focus on only thermal generators.

Let $ThermalDemand_{jt}$ be the residual demand for thermal generators after subtracting supply from wind, nuclear and other sources at weather zone j in time t.[27] I then separate $ThermalDemand$ at each zone into 12 mutually exclusive equal-frequency bins.[28] Let b_{jk} (j=1,...,8; k=1,...,12) denote the left end point of bin k for demand in weather zone j . Define

$$B_{jk}(ThermalDemand_{jt}) = \begin{cases} ThermalDemand_{jt} - b_{jk} & \text{if } ThermalDemand_{jt} > b_{jk} \\ 0 & \text{if } ThermalDemand_{jt} \leq b_{jk} \end{cases}$$

For each thermal generator i, I estimate a continuous piecewise linear model with respect to demand at each weather zone for the pre- and post-redesign periods respectively:

$$\text{Gen}_{it} = \beta_{0i} + \sum_{j=1}^{8}\sum_{k=1}^{12}\beta_{ijk}B_{jk}(ThermalDemand_{jt}) + \boldsymbol{\phi_{ih}} + \boldsymbol{\delta_{iw}} + \alpha_{1i}P_{Ng-Coal,t} + \alpha_{2i}P^{2}_{Ng-Coal,t} + \epsilon_{it} \tag{1.1}$$

I also include hour fixed effects $\boldsymbol{\phi_{ih}}$ and day-of-week fixed effects $\boldsymbol{\delta_{iw}}$, as well as a quadratic form of the price difference between coal and natural gas to control for the effect of fuel prices on the switch between coal and natural gas generation.[29] All of the coefficients are generator specific, and different before and after the market transition. Overall, there are 10,646 hourly observations in my sample. For each generator, I estimate 256 coefficients, resulting in a total of 76,032 coefficients for the 297 generating units in the sample. To form my counterfactual, I use the estimates from the pre-redesign period and calculate the change in generation quantity for each generator i at each hour t in the post-redesign period. In the mathematical form,

$$\Delta \text{Gen}_{it} = (\hat{\text{Gen}}_{it} \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{post}, \boldsymbol{X}_{\boldsymbol{t}}^{post}) - (\hat{\text{Gen}}_{it} \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}^{pre}, \boldsymbol{X}_{\boldsymbol{t}}^{post})$$

[27] Others include biomass, petroleum coke, distillate fuel oil, solar and electricity storage.

[28] The optimal number of bins is selected by using the leave-one-out cross validation technique. Specifically, given the number of bins, I estimate the corresponding model on (N-1) observations (hours) and predict the outcomes for the remaining one. I repeat the process for all N combinations and calculate the prediction errors. I experiment with different numbers of bins and choose the one that minimizes the mean squared error.

[29] Including higher-order polynomials of the fuel price difference yields similar results. I also perform ridge regressions as an alternative specification to address overfitting concern. The results are quantitatively similar.

The standard errors are estimated using the simple block wild bootstrapping method where a "block" consists of 24 hours of a calendar day. This method allows for arbitrary correlations across generators as well as serial correlations up to 24 hours.

1.6 Results

1.6.1 Effect of the Market Redesign on Generation Quantity

In this section, I first present the estimated average hourly changes in generation quantity and then examine how the changes in generation vary with changes in demand and fuel prices. In light of the heterogeneity of costs across generators, I aggregate the results according to their fuel and technology types.

Table 1.3 reports the estimated changes in generation quantity averaged over all hours in the post-redesign period. The baseline column shows the results from the estimation of equation (1.1). On average, coal generation increases by 511.2 MWh per hour while natural gas generation decreases by a similar amount. This finding suggests that the centralized market makes use of lower-cost resources more than the bilateral trading market. To provide more perspective on the magnitude of this change, the overall coal generating capacity in ERCOT is 19,819 MW. Therefore, the increase in coal generation is roughly a 3% increase in the utilization of overall coal capacity. Within natural gas generators, the decrease in generation comes from both combined-cycle and steam-turbine generators. Interestingly, combustion turbine generators experience an increase in production after the redesign, although they do not have the lowest marginal costs. This increasing usage can be explained by their ability to adjust production quickly to serve peak loads. I will discuss a more nuanced picture of these changes later.

In addition, I run several alternative specifications to check the robustness of my results. First, since ambient temperature may affect thermal generators' efficiency, I re-run my analysis including a quadratic form of temperature in model (1).[30] Second, generators differ in their ability to adjust output in response to load fluctuations. To capture this dynamic constraint, I include the daily variance of demand and one-hour lagged demand in models (2) and (3), respectively. Third, in model (4), I experiment with a higher order polynomial of the fuel price gap. Finally, in model (5), I truncate the predictions that are either below

[30]Temperature data are collected from National Centers for Environmental Information (NCEI)'s Integrated Surface Database. I use the average hourly temperature of the three largest cities in Texas, i.e., Houston, Dallas, and San Antonio.

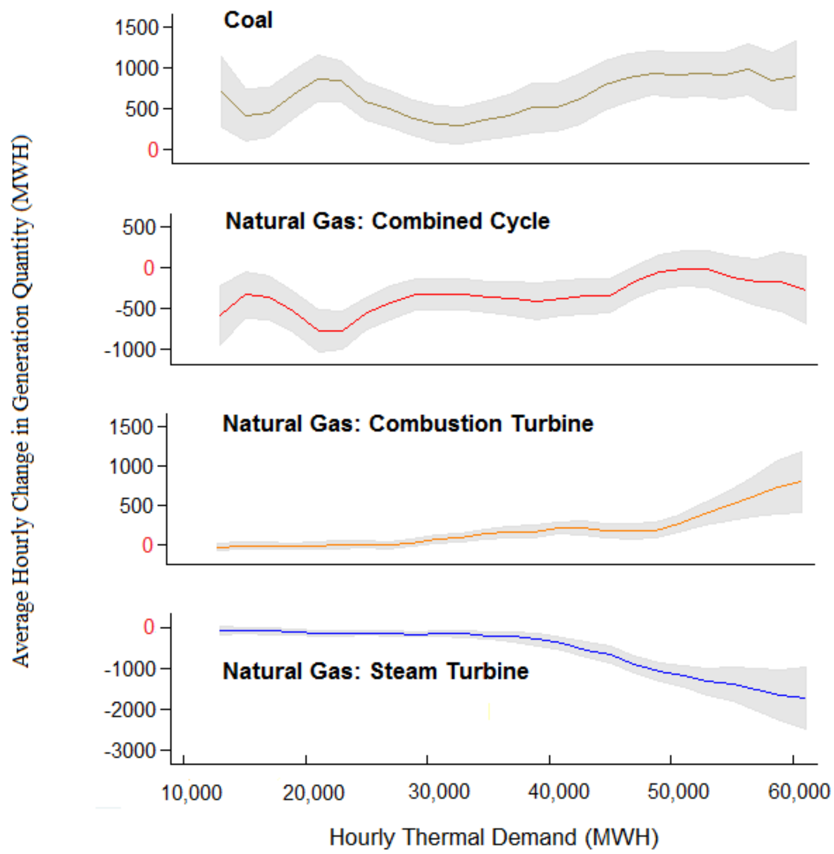
Fuel Type	Baseline Model	Alternative Models				
		(1)	(2)	(3)	(4)	(5)
Coal	511.2	524.2	508.2	509.1	554.9	480.4
	(71.2)	(68.6)	(75.6)	(73.5)	(70.8)	(71.5)
Natural Gas: CC	-353.4	-350.2	-352.7	-357.2	-362.4	-423.7
	(57.8)	(57.8)	(60.6)	(60.0)	(54.7)	(55.0)
Natural Gas: CT	106.8	86.5	106.2	113.6	115.6	95.7
	(15.4)	(13.6)	(15.2)	(16.0)	(16.5)	(15.3)
Natural Gas: ST	-306.3	-301.1	-303.3	-314.8	-281.9	-346.1
	(29.8)	(30.0)	(28.9)	(31.2)	(29.9)	(27.4)
Quadratic Fuel Price Difference	Y	Y	Y	Y	N	Y
Quartic Fuel Price Difference	N	N	N	N	Y	N
Quadratic Temperature	N	Y	N	N	N	N
Standard Deviation of Demand	N	N	Y	N	N	N
One-hour Lagged Demand	N	N	N	Y	N	N
Truncation	N	N	N	N	N	Y

Notes: This table reports the estimates of the average hourly changes in generation quantities measured in MWh, using equation (1.1) and several alternative models. For all models, hour and day-of-week fixed effects are included. The sample consists of 10,464 hourly observations and 297 generating units. Standard errors reported in the parentheses are estimated using the simple block wild bootstrapping method.

Table 1.3: Effect of the Market Redesign on Average Generation Quantities

zero or beyond the generators' nameplate capacities. Note that while about 25% of the predictions are truncated, only 5% of these are more than 5 MWh away from the thresholds. The results from these alternative specifications are reported in Table 1.3. Although the magnitudes differ slightly, the overall pattern from each alternative specification is consistent with my baseline results. In addition, in Appendix D, I perform two placebo tests and find that these changes are not observed in other years when there is no market redesign.

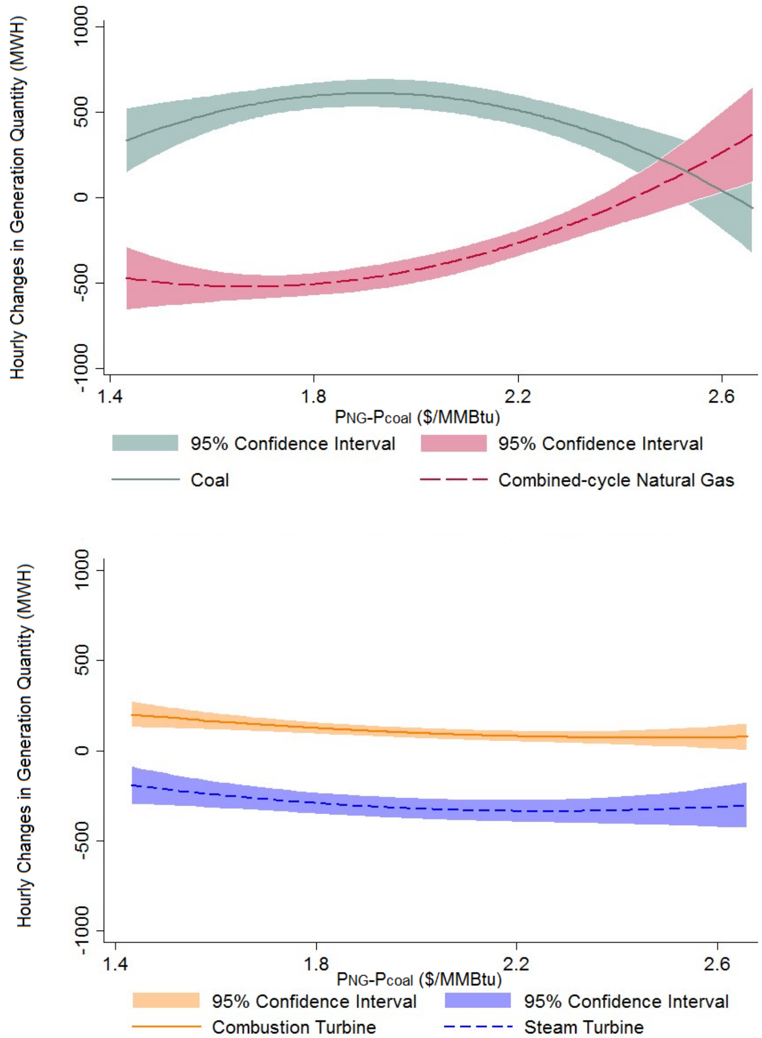
Next, I examine how the average hourly change in generation quantity varies with demand levels, assuming average post-redesign fuel prices. Figure 1.7 plots the changes for different fuel and technology types. For coal generation, the increase persists across all demand levels, but tends to be larger when demand is higher. Note that the increase is also larger when thermal demand is at about 20,000 MWh, the point at which coal and combined-cycle natural gas split on the marginal cost curve. The results show that the decrease in combined-cycle natural gas generation also changes with demand. Specifically, the decrease is greater when demand is relatively low. When demand is high, the decrease in combined-cycle natural gas generation is insignificant, as this resource is relatively cheaper compared to steam turbines.



Notes: These figures plot the average hourly changes in generation quantity at different demand levels. The prices of natural gas and coal are fixed at their post-redesign averages, i.e. $4.17 per MMBtu for natural gas and $2.17 per MMBtu for coal. The grey area indicates 95% confidence intervals.

Figure 1.7: Average Hourly Changes in Generation Quantity by Demand Levels

For combustion and steam turbines, their changes display interesting patterns. When demand is low, these two resources are less likely to be used. Consequently, the market redesign has little effect on their generation levels. However, when demand is high, combustion turbine generation increases and steam turbine generation decreases. The increase in combustion turbine generation is due to its relatively low marginal costs compared to steam turbines, as well as its ability to ramp up and down quickly, which allows it to offset steam turbine generation in those cases when coal or combined-cycle natural gas generators are unable to. Overall, these results reinforce the role of cost in determining generation outcomes in the centralized market.



Notes: This figure shows the hourly changes in generation quantity over the price differences between natural gas and coal. Demand at each zone is assumed to be at its post-redesign average, which adds up to 31,196 MW for the overall thermal demand. The shaded area indicates 95% confidence intervals.

Figure 1.8: Hourly Changes in Generation Quantity and Fuel Price Differences

Finally, in Figure 1.8, I examine the changes in generation quantity at different fuel prices, assuming average post-redesign demand levels. Regarding coal generation, the results indicate an interesting inverse-U shape. As the price gap between coal and natural gas increases, so does the switch from natural gas to coal generation. However, at a certain level, this increase in coal generation slows, as market participants can easily identify the cost advantages of coal at this gap level regardless of market design. As a result, there is less room for improvement under the centralized market. This pattern is also supported by the changes in combined-cycle natural gas generation. With marginal costs close to coal, combined-cycle generators experience changes in the opposite direction to those of coal. In contrast, the fuel price difference has a relatively small effect on combustion and steam turbine generation as these generators are further away from coal on the marginal cost curve.

1.6.2 Effect of the Market Redesign on Generation Cost

To calculate the changes in the generation cost due to the market redesign, I use each generator's estimated quantity changes from the previous section and their average marginal costs in the post-redesign period.[31] The overall change in generation cost at hour t is the sum of changes from all the generators, i.e.,

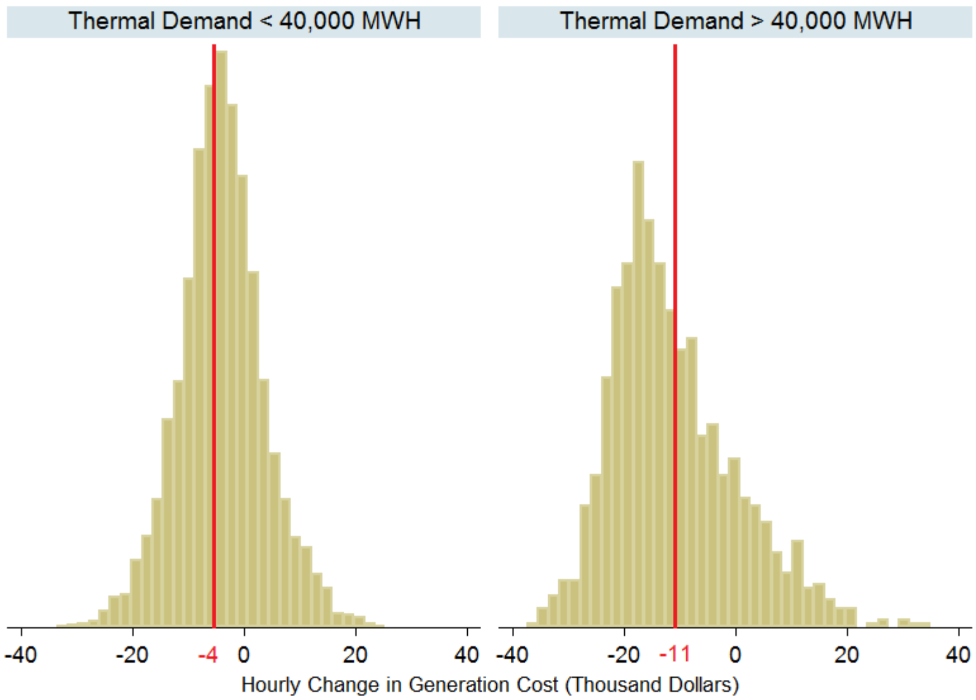
$$\Delta Cost_t = \sum_i \Delta Gen_{it} * MC_i$$

Averaging across all hours, the cost reduction is estimated to be \$5,062 per hour with a bootstrapped standard error of \$1,026 for the nine months in the post-redesign period. The reduction is roughly 0.5% of the average total hourly generation cost. Although these changes vary on an hourly basis, the generation cost is estimated to be lower than what it would have been without the market redesign for about 75% of the time. Figure 1.9 shows the distribution of hourly changes in generation cost for low and high demand hours separately. In general, we can see that the cost reduction is higher during high demand hours.

1.6.3 Effect of the Market Redesign on Emissions and External Costs

Although the centralized market leads to a significant reduction in generation costs, it also affects social welfare through changes in emissions. From a social perspective, any private

[31]This approach assumes no start-up cost and constant marginal costs. An alternative approach is to run regressions similar to equation (1.1), but with a different dependent variable – the heat inputs. In this way, I can capture any nonlinearity in the fuel usage over the operational range of a generator. With the estimated changes in heat inputs, I can apply the fuel prices to derive the changes in fuel costs. This approach, though, does not take into account other cost components, such as emission allowance costs and VOM. To address this issue, I estimate the non-fuel cost changes using the results from the original generation regressions and the non-fuel portion of each generator's marginal cost. The resulting estimate for the overall cost reduction is \$8,332 per hour on average for the nine months post redesign. This is higher than the estimate without start-up costs, as it captures the additional cost saving from increasing generation by combustion turbines which have the lowest start-up cost. The limitation of this approach is that only generators covered in CEMS data have heat input information. Thus I must limit the sample to those generators, which comprise about two thirds of all thermal generators.



Notes: This figure shows the histograms of the estimated changes in generation cost for hours of low demand (thermal demand < 40,000 MWh) and hours of high demand (thermal demand > 40,000 MWh). The red lines indicate the average change in generation cost in each case.

Figure 1.9: Distribution of Estimated Changes in Generation Cost: Low v.s. High Demand

efficiency gain must be weighted against changes in external costs of emissions.

Given the available data, I focus on three pollutants: CO_2, SO_2 and NO_x. To estimate the change in emissions of each pollutant, I again use the estimated changes in electricity generation quantity and each generator's emission rates. For pollutant j, I calculate the change in emission quantity at hour t as the sum of emission changes from all the generators, i.e.,

$$\Delta\text{Emission Quantity}_{jt} = \sum_i \Delta\text{Gen}_{it} * \text{Emission Rate}_{ij}$$

The second column in Table 1.4 reports the average hourly changes in emission quantities. On average, CO_2 emission increases by 350.5 tons per hour or 1.3 percent. This rise in CO_2 emission is not surprising given the increasing usage of coal generators in the centralized market. On average, coal power plants emit 1.16 tons of CO_2 per MWh, while natural gas plants emit only 0.65 tons of CO_2 per MWh. For SO_2 and NO_x, I find that emissions decrease by 0.268 and 0.239 tons per hour, respectively. Notably, SO_2 emissions decrease despite the fact that coal generators on average emit far more SO_2 than natural gas generators. This decrease is a result of generation changes within coal plants. The SO_2 emission rates of coal generators are dispersed, ranging from as low as 0.1 pound/MWh to as high as 14.8 pound/MWh. This variation is a reflection of heterogeneity in coal power plants' compliance strategies with environmental regulations. For example, coal plants can choose whether to install a scrubber or what kinds of coal to use. As a result, when high sulfur emitters are displaced by low sulfur emitters within coal generators, the overall SO_2 emission levels can decrease. However, this change is not significant, given the relatively large standard errors.

In a similar way, I obtain the changes in external costs associated with these emission changes. For pollutant j, the change in the external cost at hour t is calculated as follows:

$$\Delta \text{External Cost}_{jt} = \sum_i \Delta \text{Gen}_{it} * \text{Emission Rate}_{ij} * \text{Marginal Damage}_{ij}$$

To calculate the monetary value of emissions, I need to select appropriate measures of marginal damages. For CO_2, EPA (2016) compiles estimates on the social cost of carbon for use in regulatory analysis. For one metric ton of CO_2 emission in 2011, the social cost ranges from \$10 to \$51 in 2007 dollars depending on the assumed discount rates. I convert 2007 dollars to 2011 dollars to make the values comparable with generation cost estimates. For SO_2 and NO_x, I use Jaramillo and Muller (2016)'s marginal damage estimates. Unlike CO_2 which is a uniformly mixed pollutant, SO_2 and NO_x have relatively localized geographic impacts. Hence, these estimates are spatially differentiated at the county level. For SO_2, the marginal damage ranges from \$7,713.4 to \$41,959 per metric ton, while for NO_x, the marginal damage ranges from \$1,434.6 to \$7,785.2 per metric ton, both in 2011 dollars.

The rightmost column in Table 1.4 reports the average hourly changes in the external costs of the three pollutants. The external costs of CO_2 emissions range from \$3,855.5, to \$19,628 per hour depending on the social cost of carbon, while the external costs of SO_2 and NO_x emissions are estimated to be \$10,847 and \$-1,930.5 per hour respectively. Taken together, these results show that the overall change in emission costs exceeds the private generation cost savings of \$5,062 per hour. When interpreting the results, two caveats should be kept in mind. First, power plants may internalize some of the external costs due to cap and trade programs. In Texas, while CO_2 emissions are not regulated, SO_2 and NO_x are subject to cap and trade programs. However, this internalization should not

Pollutant	ΔEmission Quantity (Ton/Hour)	Marginal Damage (2011$/Ton)	ΔExternal Cost (2011$/Hour)
		11	3,855.5
			(688.6)
CO_2	350.5	35	12,267.5
	(62.6)		(2,191)
		56	19,628
			(3,505.6)
SO_2	-0.268	7,713.4 - 41,959	10,847
	(0.25)		(4,949)
NO_x	-0.239	1,434.6 - 7,785.2	-1,930.5
	(0.05)		(269.3)

Notes: This table reports the average hourly changes in emission quantities and the associated changes in external costs. Standard errors reported in the parentheses are estimated using the simple block wild bootstrapping method.

Table 1.4: Average Hourly Changes in Emission Quantities and External Costs

significantly impact my results, since the average allowance prices for SO_2 and NO_x during my sample period are just $8.4 and $275.5 per ton respectively, a very small fraction of the actual damages. A second caveat is that I implicitly assume that any environmental damage incurred by changes in SO_2 and NO_x emissions is confined to Texas. That is, no changes in emission levels are created outside of Texas as a result of this market redesign. However, if the cap and trade programs are binding, then by constraint, emission increases in Texas would create emission reductions somewhere outside of Texas, resulting in no aggregate change in emission levels. Furthermore, even without aggregate changes in emission quantity, the redistribution of the pollutants may still affect overall environmental costs, given the spatially heterogeneous nature of the marginal damages. A thorough analysis of this spillover effect is beyond the scope of this paper. Nevertheless, even without considering the impact of either SO_2 or NO_x, I find the increase in the external cost of CO_2 emissions itself completely offsets the private efficiency gain, as long as the marginal damage of CO_2 is greater than $15 per ton.

1.7 Discussion

The above results show that both the generation cost reduction and the external cost increase are statistically significant and economically large. In this section, I first compare my results with the predicted savings expected by ERCOT, and then discuss whether the redesign is warranted on a cost-benefit basis.

Prior to the implementation of the new market design, the Public Utility Commission of Texas retained several consulting firms, Tabors Caramanis & Associates (TCA) in 2004, and CRA International, Inc. and Resero Consulting (CRA/Resero) in 2008, to conduct cost-benefit analyses of the new market design. These studies use the GE MAPS simulation model that includes a full transmission representation of ERCOT but assumes no market power. The annual production cost reduction is estimated to be \$66.8 million and \$48.0 million in 2008 dollars, respectively. Furthermore, a back-cast using bids submitted during a market trial suggests that the 2008 production cost would have been lower by \$90 to \$180 million, had the centralized market design been in place at that time (ERCOT, 2011a). In the previous section, I find that the average hourly cost saving is \$5,062, which amounts to \$44.3 million on an annual base. This number is smaller than the aforementioned estimates, but of the same order of magnitude. The difference between my estimate and the engineering estimates suggests that changes in market power during the redesign is also important in determining generation cost savings.

Benefit	
Generation Cost Saving(\$m/year)	44.3 (Author's Calculation)
Ancillary Services Cost Saving(\$m/year)	17.0 (ERCOT, 2011a)
Savings from Improved Generation Siting(\$m/year)	34.9 (CRA/Resero, 2008)
Cost	
One-time Implementation Cost(\$m)	548.6 (ERCOT, 2011b)
Incremental Operational Costs(\$m/year)	14.6 (CRA/Resero, 2008)
Environmental Cost(\$m/year)	111.9-250.0 (Author's Calculation)

Notes: This table lists the benefits and costs of ERCOT's market redesign. All numbers are in 2011 dollars.

Table 1.5: The Cost-Benefit Analysis of ERCOT's Market Redesign

A second question that arises is whether the market redesign is warranted when costs and benefits are considered. The transition to the centralized market provides a number of benefits in addition to the decrease in generation cost. Specifically, the centralized market is expected to reduce annual ancillary service costs by \$17 million per year.[32] Additionally, in the long run, the centralized market can lead to improvement in siting of new resources through more transparent locational marginal prices. CRA/Resero (2008) estimates this benefit to be \$34.9 million per year. However, this transition also carries several costs, principally the external costs from increasing emission levels, but also a one-time implementation cost of \$ 548.6 million as well as yearly recurring expenses of \$14 million. Table 1.5 summarizes these benefit and cost components. Taken together, the picture that emerges is that the redesign will be cost effective for the first 10 years of operation if the discount rate is less than 8% without considering environmental costs. However, the market redesign

[32]Ancillary services are those services necessary to maintain grid stability and support continuous balance between supply and demand.

will create a social welfare loss when the environmental impacts are taken into account.[33]

1.8 Conclusion

This paper examines the impact of the Texas electricity market redesign on both market efficiency and social welfare. To do so, I use a flexible semi-parametric approach to estimate the changes in generation allocation among different types of generators. I then use these estimates to quantify the associated changes in production costs and emissions. My results show that the market redesign improves market efficiency, suggesting that the informational benefits created by a centralized market outweighs any change in market power incentives. Currently, a centralized market design is the norm for all deregulated electricity markets in the US. This paper provides evidence supporting such a practice on the basis of efficiency. Worldwide, there are still regions that either have not restructured their electricity markets or have adopted a bilateral trading model. The Texas' experience also provides a useful reference for those regions that may consider a move to a centralized market. Given that my estimates are based on the market conditions in Texas, in future research, it would be interesting to conduct a cross-market comparison to better understand any market-specific drivers that may impact the direction and magnitude of the efficiency changes.

While my results attest to the superiority of the centralized market design in terms of efficiency, I also find that the transition to a centralized market increases emission levels. The conflict between efficiency improvement and pollution mitigation is a result of the disparity between private and social costs, rather than flaws in the design per se. Setting up carbon pricing schemes or appropriate emission caps provides one solution to resolve this conflict. In recent years, the idea of an ISO-administered "carbon adder" – a price on carbon added to generators' bids – has been proposed as an alternative way to reduce carbon emissions. I find the environmental cost associated with increasing emissions is not trivial. In fact, when those costs are taken into account, the redesign no longer passes a cost-benefit test. This finding highlights the need to take environmental impact into account when we make decisions in the energy market.

[33] One caveat that should be noted is that the extrapolation is based on the market conditions between June 1, 2010 to August 31, 2011. During the subsequent years, natural gas prices have dropped from about $4/ MMBtu to less than $2/ MMBtu. This decrease in natural gas prices has led to widespread substitution of natural gas for coal (Cullen and Mansur, 2014). On the one hand, since some natural gas power plants, especially the combined-cycle plants, are ahead of coal in the merit order, we may not see as much displacement of natural gas generation by coal generation in the later years as in 2011. Hence, the environmental cost may be substantially lower in the later years than the estimates shown here. On the other hand, private cost savings from the market redesign may also be lower, given smaller differences in marginal costs among generators.

CHAPTER 2

Greener Fuel, Bluer Sky? The Impact of Motor Fuel Standards on Air Quality in China

2.1 Introduction

Air quality in China is notoriously poor. As one example, on January 12, 2013, Beijing saw a jaw-dropping reading of 755 on the Air Quality Index (AQI) which nominally maxes out at 500.[1] According to the World Bank, 16 Chinese cities appeared on the list of the world's top 20 most polluted cities.[2] Asia Development Bank also reports that less than 1 percent of the largest Chinese cities meet the air quality standards recommended by the World Health Organization (Zhang and Crooks, 2012).

The transportation sector is a major source of air pollution in metropolitan areas of China. For example, car emissions account for 22.2 percent of Beijing's $PM_{2.5}$ particles, according to the monitoring data released by the Beijing Municipal Environmental Protection Bureau in 2012.[3] Motor vehicles are also a significant source of CO, HC, and NO_x emissions, all of which are produced through inefficient or incomplete combustion. Fuel quality is a key determinant of vehicle emissions. It is now understood that lowering fuel sulfur levels not only reduces direct emissions of various pollutants, but also enables the use of certain advanced pollution control technologies in vehicles. In fact, sulfur levels must be reduced to near zero if the maximum benefits are to be achieved by the most advanced technologies used today (Walsh, 2013).

In recognition of the importance of low-sulfur fuel, China, among many other coun-

[1]The reading is recorded by an air-quality monitoring device atop the United States embassy in Beijing.
[2]"The Most Polluted Places On Earth". CBS News. January 8, 2010 (Retrieved March 23, 2014).
[3]http://www.china.org.cn/environment/2013-09/25/content_30129133.htm (Retrieved March 23, 2014)

tries, set up a series of fuel quality standards to gradually reduce the sulfur content levels. The introduction of CHINA III fuels is a major upgrade in the recent decade. The maximum sulfur level is mandated to be 350 ppm lower for gasoline and 1650 ppm lower for diesel. Despite the drastic change in sulfur content levels, little is known about its overall effectiveness in reducing air pollution. In the United States and Europe, some research programs were established to estimate the relationship between reductions in sulfur levels and changes in vehicle emissions.[4] However, the results from these engineering models depend heavily on the assumed on-road vehicle fleet and the catalytic systems in use, which renders their estimates not directly applicable to the Chinese context. Furthermore, as established by many economic studies, behavioral responses from both consumers and producers are important factors in determining the actual effectiveness of an environmental policy. On the consumer's side, Davis (2008) shows that the driving restriction in Mexico City is ineffective in reducing air pollution because it led to an increase in the total number of vehicles in circulation as well as a change in composition toward high-emission vehicles. On the supplier's side, Auffhammer and Kellogg (2011) finds that the federal gasoline regulation standards, which allow refiners' flexibility in choosing a compliance mechanism did not improve air quality because the compliance options chosen by refiners did not reduce emissions of those compounds most prone to forming ozone. An engineering model cannot possibly take into account such behavioral responses.

This paper seeks to fill this gap by assessing the impact of CHINA III fuel standards on air pollution using actual daily air quality data. Air pollution levels are compared within a relatively narrow time window before and after the implementation of the new standards. The analysis controls for confounding factors by employing a regression discontinuity (RD) design and including a rich set of explanatory variables. In the primary specification, I use high-order polynomial time trends to control for time-varying unobservables. As a robustness check, I also consider a local linear regression using only observations within a two-year window. To evaluate the credibility of the identification strategy, I complement the RD analysis with a difference-in-difference method (DID) approach, taking advantage of the spatial and temporal variation with which the regulations were applied. These approaches are similar to Auffhammer and Kellogg (2011) and Chen and Whalley (2012).

The primary finding of this paper is that while the introduction of CHINA III gasoline

[4]One such program is the European Program on Emissions, Fuels and Engine Technologies (EPEFE). It finds that exhaust emissions of HC, CO and NO_x went down as the fuel sulfur level dropped from 382 ppm to 18 ppm. The effects were generally linear at around 8-10% in urban driving and 20-50% in high speed driving (Petit et al, 1996).

standard substantially reduces air pollution, there is no evidence of improvement as a result of CHINA III diesel standard. Air pollution is reduced by 6% on average after CHINA III gasoline standard came into effect. By contrast, the average reduction effect of CHINA III diesel standard is only 1.5% and insignificant. These results are robust to alternative specifications and measures. I then explore potential explanations for these divergent outcomes and find two contributing factors: the lack of cost pass-though and the presence of a loophole in the diesel policy. First, desulfurization is more costly for diesel than for gasoline. But refineries are not able to pass the incremental costs to final consumers due to retail price caps. The lack of price differentials discourages refineries from installing costly desulfurization facilities. While such non-compliant motives exist for both gasoline and diesel, the presence of a loophole in the diesel policy makes evasion only possible for diesel. Specifically, off-road diesel is subject to a less stringent standard but indistinguishable from on-road diesel to naked eyes. Hence, off-road diesel may be used to fill on-road vehicles without being detected at gas stations. Given its illegality, irrefutable evidence is difficult to obtain, but anecdotal evidence from news articles confirms such a practice.

With the pollution reduction estimates, I also examine the magnitude of the associated health benefits in economic terms. I quantify the annual health benefits of CHINA III gasoline standard by using the well documented mortality effect of PM_{10} exposure in the epidemiological literature. An annual reduction of PM_{10} by 7.7 percent leads to an estimate of 23,665 lives saved nationwide. Using the value of a statistical life suggested by Wang and He (2010), I find the economic benefit amounts to \$3.0 billion per year, outweighing the upgrading cost of \$28.2 million. This result justifies promoting low-sulfur gasoline on a cost-effective basis. This estimate also serves as a lower bound for the benefit of CHINA III diesel standard that is foregone.

The remainder of the chapter goes as follows: Section 2.2 and 2.3 provides a description of the policy background and the data I use. Section 2.4 introduces the identification strategies and presents the results from both main and alternative specifications. Section 2.5 discusses the implication of the results, and Section 2.6 concludes.

2.2 Policy Background

Sulfur is a natural component in crude oil that ends up in gasoline and diesel unless being removed. Fuel sulfur not only contributes to the sulfur dioxide levels in ambient air, but also undermines the effectiveness of catalytic converters and leads to higher tailpipe emissions of other pollutants such as nitrogen oxides, carbon monoxide and fine particles. Upon combustion, fuel sulfur is oxidized to sulfur oxides, primarily sulfur dioxide (SO_2) with small amounts of sulfur trioxide (SO_3). Sulfur oxides combine easily with base metal oxides and form sulfates which constitute a significant portion of the particle mass in the atmosphere. More importantly, the presence of sulfur as an oxide or sulfide inhibit the catalytic function of automobile exhaust catalysts. The sulfur species adsorb on the catalyst site which is then not available for the preferred catalytic reactions, resulting in less overall activity and more emissions (Manufacturers of Emission Controls Association, 1998).

In recognition of the key role sulfur plays in reducing vehicle emissions, the Environmental Protection Agency (EPA) in the US has adopted increasingly stringent standards on sulfur content over the past decade. Under Tier 2 Rule, all gasoline sold in the US must meet an average sulfur level of 30 ppm with a 80 ppm cap starting from 2006. A 15 ppm sulfur specification, known as ultra low sulfur diesel (ULSD), was phased in for highway diesel from 2006 to 2010 and for nonroad, locomotive, and marine (NRLM) engines from 2007 to 2014. These fuel requirements, coupled with advanced emission control technologies, were expected to decrease emissions by more than 90%. In 2013, the EPA proposed Tier 3 Standard which further tightens the cap to 10 ppm.[5] In Europe, fuel sulfur levels have been declining as well, since at least 1980. The sulfur level for gasoline and highway diesel was reduced to a uniform standard of 50 ppm in 2005. The standards are currently set at 10 ppm maximum for all transportation fuels.

By contrast, in China, progression to low-sulfur fuel was very slow. Despite the mounting health consequences of severe air pollution, the first national fuel sulfur regulation did not come through until the early 2000. The maximum sulfur levels were set at 500 ppm for gasoline and 2000 ppm for diesel, respectively. The standards remained unchanged until 2010. Starting from January 1, 2010, gasoline must meet CHINA III standard with a 150 ppm cap. On July 1, 2011, diesel must meet CHINA III standard as well. Table 2.1 outlines major steps China has taken or plan to take to improve motor fuel quality.[6] The standards generally follow European precedents, with minor adjustment. The current national standard for gasoline and diesel is CHINA IV.

This paper focuses on the introduction of CHINA III standards for gasoline and diesel.

[5]For more information about fuel sulfur regulations in the US, please refer to the EPA's website: http://www.epa.gov/otaq/fuels/index.htm.

[6]Although these standards also regulate other chemical contents, the predominant change in each standard is the maximum sulfur level.

	Gasoline			Diesel		
Stage	Standard	Maximum Sulfur Level (ppm)	Implementation Date	Standard	Maximum Sulfur Level (ppm)	Implementation Date
China I	GB 17930-1999	lead-free		GB 252-2000	2000	01/01/2002
China II	GB 17930-2004	500	07/01/2005	GB/T 19147-2003	500 (voluntary)	10/01/2003
China III	GB 17930-2006	150	01/01/2010	GB 19147-2009	350	07/01/2011
China IV	GB 17930-2011	50	01/01/2014	GB 19147-2013	50	01/01/2015
China V	GB 17930-2013	10	01/01/2018	GB 19147-2013	10	01/01/2018

Table 2.1: Major Steps of China's Nationwide Fuel Quality Improvement

There are two reasons to focus on CHINA III. First, the magnitude of this upgrade is the largest in the recent decade. The implementation of CHINA III is set to reduce the maximum sulfur level by 350 ppm for gasoline and 1650 ppm for diesel.[7] Second, the availability of consistent air quality and weather measures during that period allows me to compare and identify the changes in air quality due to the new standards.

Cities in China may choose to develop and implement their own fuel quality standards without requiring national-level approval. The standards are usually issued by local environmental protection agency after negotiating with the major state-owned oil companies. The four most developed cities in China, namely Beijing, Shanghai, Guangzhou and Shenzhen, promoted low-sulfur fuels earlier than the national implementation dates. Table 2.2 summarizes the implementation dates for those cities. One motivating factor to implement higher fuel quality standards is to prepare for international events. For example, Beijing and Shanghai rolled out CHINA IV standards about 6-8 months prior to the Olympics and the World Exposition. Such variations in timing allow me to do a difference-in-difference analysis as a secondary specification.

[7] CHINA II diesel is only a recommended standard. It has never been enforced.

City	Gasoline			Diesel		
	China III	China IV	China V	China III	China IV	China V
Beijing	07/01/2005	01/01/2008	08/01/2012	07/01/2005	01/01/2008	08/01/2012
Guangzhou	05/01/2008	08/01/2010		05/01/2008	01/01/2014	
Shanghai	–	10/01/2009	12/01/2013	–	10/01/2009	12/01/2013
Shenzhen	03/01/2007	01/01/2011		03/01/2007		
Nation	01/01/2010	01/01/2014	01/01/2018	07/01/2011	01/01/2015	01/01/2018

Table 2.2: Summary of Implementation Dates

2.3 Data

The sample period includes all observations within a 6 year window around the national implementation date of China III gasoline, which runs from January 1, 2007 through December 31, 2012. It is the largest symmetric window in which air pollution indexes are available. This section provides more details on the data I use.

2.3.1 Air Pollution Index (API)

Air quality is measured and recorded by the Ministry of Environment Protection (MEP) of China. On June 5, 2000, MEP started to publish a daily Air Pollution Index (API) for 86 cities including all provincial capitals and municipalities. MEP continued to publish API for those cities until January 15, 2013 when the new Ambient Air Quality Standard (AAQS) came into effect. Under AAQS, MEP reports a new measure, namely Air Quality Index (AQI), for 58 cities (including all provincial capitals and municipalities), leaving the rest covered in the original API calculation.

For each city, MEP measures the concentrations of three pollutants, i.e., NO_2, SO_2 and PM_{10}, and converts daily mean into a pollutant-specific API ranging from 0 to 500.[8] The pollutant-specific API is defined as a linear interpolation between two breakpoints such that

$$API_x = \frac{C_x - C_{x,l}}{C_{x,u} - C_{x,l}}(I_{x,u} - I_{x,l}) + I_{x,l} \quad x \in \{SO_2, NO_2, PM_{10}\} \tag{2.1}$$

[8] The description here is based on Technical Requirements for Urban Air Quality Daily Report ([2000] No. 26). MEP stipulates the number of stations according to city population and size, and computes the daily city averages over all stations and all hours. The daily report uses data from 12:00pm from the previous day to 12:00pm of today.

Daily Average concentration ($\mu g/m^3$)				Air Quality Grade	Air Quality Conditions	Health Implications
API	PM_{10}	SO_2	NO_2			
500	600	2620	940	V	Heavily Polluted	Healthy people will experience reduced endurance in activities. There may be strong irritations and symptoms and may trigger other illnesses. Elders and the sick should remain indoors and avoid exercise. Healthy individuals should avoid outdoor activities.
400	500	2100	750			
300	420	1600	565	IV	Moderately Polluted	Healthy people will be noticeably affected. People with breathing or heart problems will experience reduced endurance in activities. These individuals and elders should remain indoors and restrict activities.
200	350	800	280	III	Slightly Polluted	Slight irritations may occur for susceptible population. Individuals with breathing or heart problems should reduce outdoor exercise.
100	150	150	120	II	Good	No health implications
50	50	50	80	I	Excellent	

Source: Technical Requirements for Urban Air Quality Daily (2000), China

Table 2.3: API Cutoff Points and Corresponding Health Implications

Where API_x is the air pollution index for pollutant x. $C_{x,l}$ and $C_{x,u}$ are the lower and upper boundaries that the concentration of pollutant x C_x falls in, and $I_{x,u}$ and $I_{x,l}$ are the corresponding breakpoints of the overall API.[9] The cutoff points for different levels of API are summarized in Table 2.3. The overall API is the maximum of all pollutant-specific APIs. If that maximum exceeds 500, the overall API is capped at 500. MEP partitions the overall API into five categories: 0-50, 50-100, 100-200, 200-300, and 300-500, and provides health recommendations for each grade. In addition, the primary pollutant is reported when API is greater than 50. Therefore, the concentration of the primary pollutant can be inferred from the API score when the primary pollutant is identified. Table 2.4 reports the summary statistics. In the sample, 69.53 % of observations reported PM_{10} as the primary pollutant, 0.14 % reported NO_2 and 6.30% reported SO_2. The remaining 24.03 % have an API below 50.

2.3.2 Meteorological Data

Meteorological factors are correlated with air pollution. I use the meteorological data obtained from Global Summary of the Day (GSOD). This dataset come from the National Climatic Data Center (NCDC) under the National Oceanic and Atmospheric Administration (NOAA) of the United States. The daily weather summaries are based on data exchanged under the World Meteorological Organization (WMO) World Weather Watch Program, which are ultimately collected by the weather stations under the China Meteorological Administration (CMA).

Of the 86 cities that are covered in API reports, 59 cities have daily weather records during the sample period. Figure 2.1 shows the locations of these cities. The variables reported are mean temperature (TEMP, .1 Fahrenheit), maximum temperature (MAX, .1 Fahrenheit), minimum temperature (MIN, .1 Fahrenheit), mean wind speed (WDSP, .1 knots), maximum sustained wind speed (MXSPD, .1 knots), precipitation amount (PCP, .01 inches) and mean visibility (VSB, .1 miles). Their summary statistics are also reported in Table 2.4.

[9] For example, if the daily mean of SO_2 is 600 $\mu g/m^3$, the corresponding API for SO_2 is (600-150)/(800-150)*(200-100)+100=169.

Figure 2.1: Location Map of the Sample Cities

2.4 Empirical Strategies and Results

In this section, I introduce the empirical approaches and present the results. The goal is to identify the extent to which fuel quality improvement ameliorates air pollution. Following Davis (2008), Auffhammer and Kellogg (2011) and Chen and Whalley (2012), I assess the impacts by employing both a time series regression discontinuity (RD) design and a difference-in-difference method (DID). In the primary specification, I conduct a time series regression discontinuity (RD) with high-order polynomial time trends using the full sample. I also consider two alternative specifications: a local linear regression using only observations within a two-year window and a difference-in-difference method (DID) taking advantage of the spatial and temporal variation with which the regulations were applied.

2.4.1 Main Results: Regression Discontinuity (RD) Design

The idea of this approach is to focus on a narrow window of time around the implementation of new fuel standards so that observations before the change provide a comparison group for observations after the change. The usual omitted variable problem does not con-

Variable	N	Mean	Std	Min	Max
API	127,808	67.76	30.94	0	500
Primary Pollutant					
PM_{10}	88,870	103.90	50.38	52	598.4
NO_2	173	118.19	21.46	81.6	198.4
SO_2	8,056	97.06	43.07	52	520.5
Not Reported	30,709				
Grade					
I		0.2403			
II		0.6750			
III		0.0808			
IV		0.0021			
V		0.0019			
Weather					
TEMP	127,808	58.35	20.61	-23.3	96.9
MAX	127,808	66.96	20.62	-14.3	110.3
MIN	127,808	49.88	21.59	-31	89.2
WDSP	127,808	4.89	2.60	0	34.4
MXSPD	127,776	9.11	4.51	1.9	46.6
PCP	123,781	0.11	0.40	0	17.36
VSB	127,804	8.08	4.58	0	22.1

Table 2.4: Descriptive Statistics of the Data

taminate identification, as long as the unobserved factors do not change discontinuously when new fuel standards are phased in (Hahn et al, 2001).

$$\ln(API_{ct}) = \boldsymbol{\alpha_c} \cdot \boldsymbol{Treat_{ct}} + \boldsymbol{\beta_c} \cdot \boldsymbol{W_{ct}} + \boldsymbol{\lambda_c} \cdot \boldsymbol{D_t} + f_c(t) + \mu_c + \varepsilon_{ct} \tag{2.2}$$

The estimation equation is given in (2.2). $Treat_{ct}$ here is a vector indicating the implementation status of CHINA III for gasoline and diesel. CHINA II gasoline and CHINA I diesel form the baseline against which these standards are compared.[10] All the coefficients are city-specific. In other words, I estimate (2.2) one city at a time. I exclude Beijing, Shanghai, Guangzhou and Shenzhen from the sample because they implemented CHINA III standards for gasoline and diesel at the same time, making it impossible to separate the effects. W_{ct} is a set of weather controls including current and one-day lag quartic polynomials in mean, max and min temperatures, mean wind speed, maximum sustained wind speed, precipitation as well as the interaction between mean temperature, mean wind speed and precipitation. D_t denotes other controls including year fixed effects, month fixed effects,

[10] Note that CHINA II diesel is only a recommended standard that has never been enforced.

day-of-week variables, and indicators for national holidays and heating days.[11] The vector $f_c(t)$ is a 9th-order Chebychev polynomial time trend that flexibly controls for time-series variation in pollution that would have occurred in absence of fuel quality upgrades. Fuel at retail stations and gas tanks does not completely turn over in a single day. I therefore allow for a 2-month linear phase-in of the treatment effect after the official implementation dates.[12] Finally, μ_c are city fixed effects. The standard errors are clustered on each month-year combination to allow for serial correlation within a month.

The first two columns in Table 2.5 report the results from estimations of equation (2.2). The mean effect of CHINA III gasoline is -0.060 with p-value 0.054. Therefore we can reject the mean effect is zero at close to the 95% confidence level. By contrast, the mean effect of CHINA III diesel is only -0.015 with p-value 0.664. We cannot conclude that air quality has been significantly improved after CHINA III diesel standard came into effect. The difference is also illustrated by the distribution of the estimates shown in Figure 2.2. The distribution of the gasoline estimates is shifted to the left relative to that of diesel. The Kolmogorov-Smirnov test for equality of distributions returns a p-value of 0.09, which suggests that we can reject that the distributions of gasoline and diesel estimates are the same at the 90% confidence level. Figure 2.3 further plots the daily log API residuals for the city of Hangzhou that is characteristic of the results. The fitted line is the time series of predicted values of the treatment effects plus the polynomial time trend. The points plotted are the sum of this fitted line with the residuals from the regression. We can see a clear reduction in the air pollution level after CHINA III gasoline is fully phased in. By comparison, the RD effect of CHINA III diesel is hardly discernable.

Because API is an aggregated index used only in China, it is difficult to interpret and compare the results with the findings of previous studies. About 70 percent of observations in the sample reported PM_{10} as the primary pollutant. Given the way API is constructed, I am able to back out the level of PM_{10} from the aggregated API for those observations. The third and fourth columns in Table 2.5 show the RD results using log PM_{10} as the dependent variable. The average effect of CHINA III gasoline on log PM_{10} is -0.077, while the average effect of CHINA III diesel is 0.013. Again, the p-values suggest that only the average

[11] The indicator for heating days takes on value one for cities north of Huai River and Qinling Mountains between November 15 to March 15, and zero otherwise. Almond et al (2009) shows that the heating policy in China leads to higher pollution concentrations in northern China where coal is burned for winter heating of homes and offices.

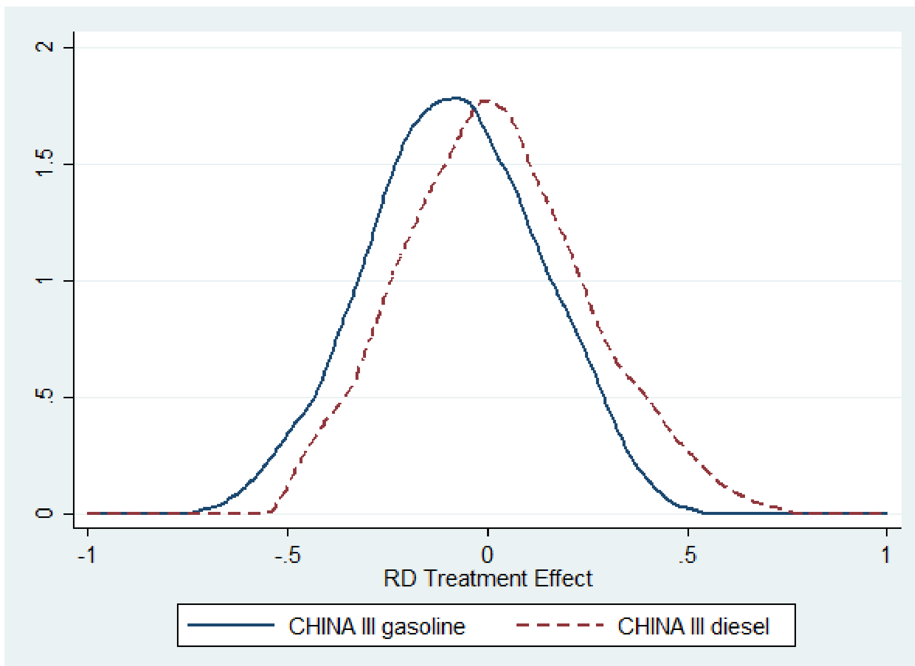
[12] According to practitioners, complete natural displacement requires 2 to 3 turnovers at gas stations and oil depots. The process generally takes 1-2 months, depending on the sales amount.

effect of CHINA III gasoline is significantly different from zero.

	Effect on ln(API)		Effect on ln(PM_{10})		Effect on ln (VSB)	
	Gasoline	Diesel	Gasoline	Diesel	Gasoline	Diesel
Mean	-0.060	-0.015	-0.077	0.013	0.127	0.004
P-value	0.054	0.664	0.023	0.732	0.000	0.878

Notes: Estimates shown are average effects of CHINA III gasoline and diesel standards across cities using the regression discontinuity specification (2.2). Standard errors are clustered at month-year level. The P-values are obtained from the test of the null hypothesis H_0: the mean effect is zero against the alternative hypothesis H_a: the mean effect is not zero.

Table 2.5: Regression Discontinuity Results: Six-year Window



Notes: The figure displays the smoothed cross-city distribution of estimated treatment effects from equation (2.2). The smoother uses an Epanechnikov kernel with a bandwidth of 0.1.

Figure 2.2: Distribution of Estimated RD Effects: Primary Specification

2.4.2 Robustness Checks

In this section, I conduct several robustness checks with different covariates and dependent variables. The first exercise is to experiment with different orders of polynomials for time trends. The top three rows in Table 2.6 report average effects and p-values from regressions

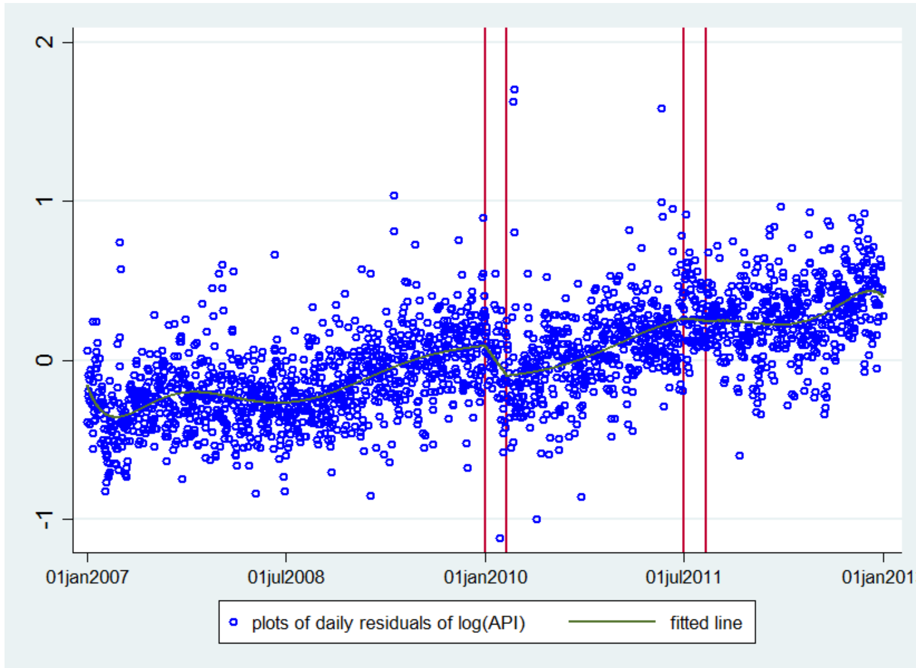


Figure 2.3: Daily Log API Residuals for the City of Hangzhou

of model (2.2) with seventh-, eighth- and tenth-order of polynomials. The estimates from these specifications are consistent with the main results, suggesting that on average, the change of gasoline standard has a more salient effect on air quality than that of diesel.

Second, I examine the changes in gasoline and diesel prices as potential confounding factors. Retail fuel prices are heavily regulated in China, as policymakers are very concerned about its impact on inflation. The National Development and Reform Commission (NDRC) sets price caps for retail gasoline and diesel based on crude prices in the global market.[13] Figure 2.4 shows the retail fuel price caps from 2007 to 2012. It generally follows an upward trend. The two vertical lines indicate the time when the new fuel standards came into effect. There were no changes during the time when CHINA III gasoline and diesel standards were phased in. This lifts the concern that the RD results are contaminated by the adjustment in driving behaviors in response to fuel price changes. Nevertheless, I include gasoline and diesel price caps in model (2.2) and redo the analysis for each city. The results shown in the fourth row of Table 2.6 suggest that the mean effects are slightly smaller, but still exhibit the same pattern.

[13] The current pricing system adjusts retail gasoline and diesel prices every 10 working days if price changes in international oil markets (Brent, Dubai and Cinta) are more than 50 CNY per metric ton. Prior to that, domestic fuel prices were adjusted when prices for crude moved by more than 4 percent over 22 working days.

	CHINA III gasoline		CHINA III diesel	
	Mean Effect	P-value	Mean Effect	P-value
Seventh-order polynomial time trend	-0.087	0.008	0.027	0.311
Eighth-order polynomial time trend	-0.065	0.035	-0.027	0.425
Tenth-order polynomial time trend	-0.067	0.034	-0.011	0.756
Including fuel prices	-0.056	0.078	-0.015	0.669
Local linear regression	-0.076	0.000	-0.016	0.488

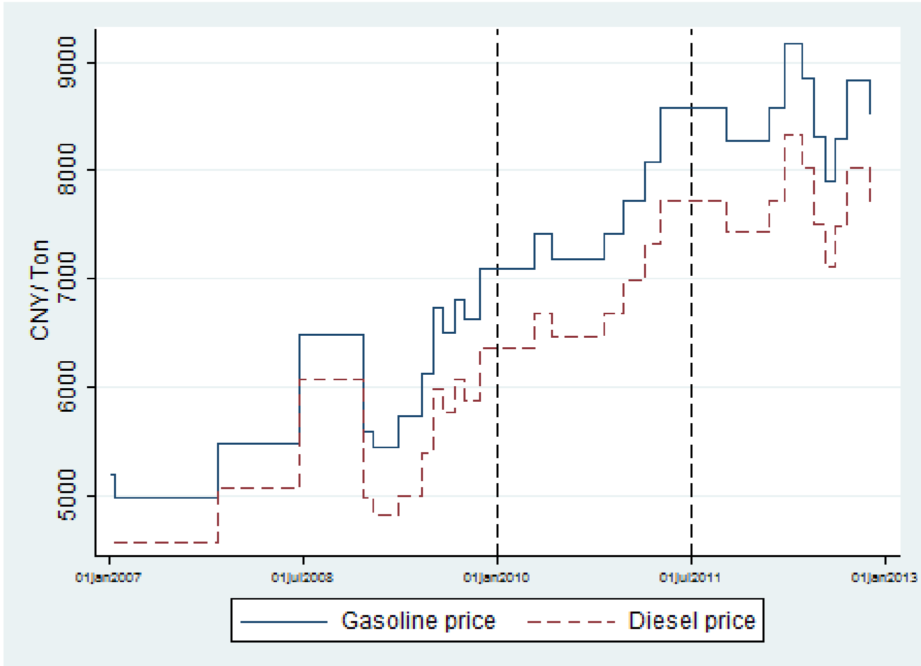
Notes: Estimates shown are average effects of CHINA III gasoline and diesel standards across cities using alternative specifications with different time trends, time windows and control variables. The P-values are obtained from the test of H_0: the mean effect is zero vs. H_a: the mean effect is not zero.

Table 2.6: Regression Discontinuity Results: Robustness Check

Finally, I address the concern about the credibility of API measures. Andrews (2008), Chen et al (2012) and Ghanem and Zhang (2013) provide suggestive evidence that some Chinese cities manipulate their air pollution data in response to the incentives set by the central government. They find that some cities reported dubious pollution data that leads to discontinuity at API equal to 100 which is the cutoff for "blue-sky days." To resolve the issue with the credibility of air quality measures, I use visibility as a proxy for air quality. Visibility is defined as the greatest distance at which a black object of suitable dimensions, situated near the ground, can be seen and recognized when observed against a bright background. Numerous studies have demonstrated that particulate matters can cause visibility impairment (Sisler and Malm, 1994). Visibility data are routinely collected at meteorology stations throughout the world. Since weather stations are less prone to political interference, visibility can be used as an alternative measure of air quality. The last two columns in Table 2.5 report RD estimates using log visibility as the dependent variable. The RD estimates suggest that the average effect of CHINA III gasoline is 0.127, while the average effect of CHINA III diesel is 0.004. Positive estimates indicate greater visibility, hence better air condition. Again, only the effect of gasoline is significant, which is consistent with the previous findings.

2.4.3 Alternative Specification: Local Linear Regression Discontinuity Design

In this section, I consider an alternative specification that only uses observations within a two-year window around the implementation dates as an important validity check. Generally, choosing a sample containing more observations increases the preciseness of the



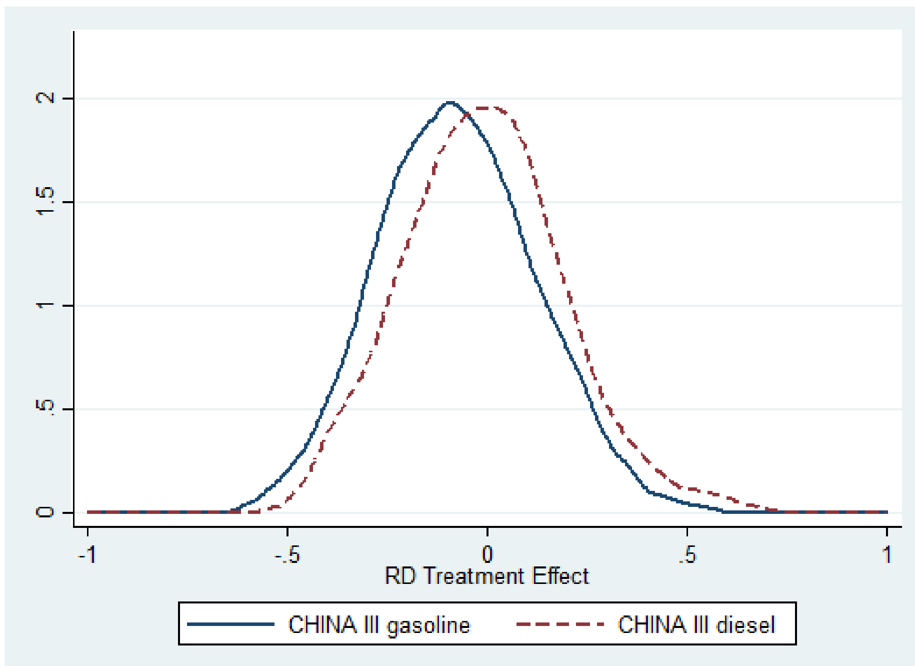
Notes: The vertical lines indicate the dates when CHINA III gasoline and diesel standards took effect.
Data Source: National Development and Reform Commission of China

Figure 2.4: Retail Price Caps in China: 2007-2012

estimates, but also causes concerns, as those observations further away from the implementation threshold are less comparable to each other. A two-year window seems to balance this trade-off by allowing to control for full seasonality with a minimal number of years involved. Following Imbens and Lemieux (2008), I fit linear regression functions to the observations within one year on either side of the phase-in period. The estimation equation is given in (2.3). $Treat_{ct}$ here is a variable indicating the implementation status of CHINA III. I estimate this equation separately for gasoline and diesel using observations within a two-year window centered around each phase-in period. Time trends are allowed to differ before and after.

$$\ln(API_{ct}) = \alpha_c \cdot Treat_{ct} + \boldsymbol{\beta_c} \cdot \boldsymbol{W_{ct}} + \boldsymbol{\lambda_c} \cdot \boldsymbol{D_t} + \gamma_c \cdot t + \delta_c \cdot t \times Treat_{ct} + \mu_c + \varepsilon_{ct} \tag{2.3}$$

The last row in Table 2.6 reports the local linear RD results. The mean effect of CHINA III gasoline is -0.076 with p-value 0.000, while the mean effect of CHINA III diesel is only -0.016 with p-value 0.488. The difference is also illustrated by the distribution of the estimates shown in Figure 2.5. The Kolmogorov-Smirnov test returns a p-value of 0.03, suggesting that we can reject the equality of distributions at the 95% confidence level. Similar to the previous findings, only CHINA III gasoline standard has significantly improved air quality. Figure 2.6 and 2.7 also plot the daily log API residuals for cities that are characteristic of these results. Figure 2.6 depicts the RD result of CHINA III gasoline for the city of Guilin. There is a noticeable reduction in the air pollution level after CHINA III gasoline is fully phased in. Accordingly, the point estimate is -0.077. By comparison, the RD effect of CHINA III diesel for the city of Hangzhou is hardly discernable, as shown in Figure 2.7. The point estimate suggests that the new diesel standard only reduced API by 2 percent.



Notes: The figure displays the smoothed cross-city distribution of estimated treatment effects from equation (2.3). The smoother uses an Epanechnikov kernel with a bandwidth of 0.1.

Figure 2.5: Distribution of Estimated RD Effects: Local Linear Regression

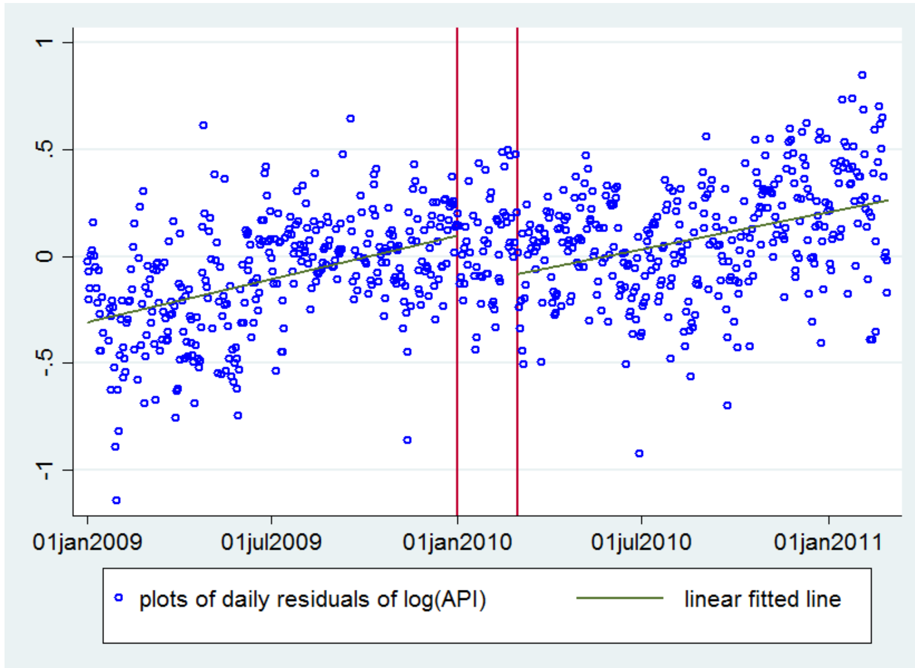


Figure 2.6: Daily Log API Residuals for the City of Guilin

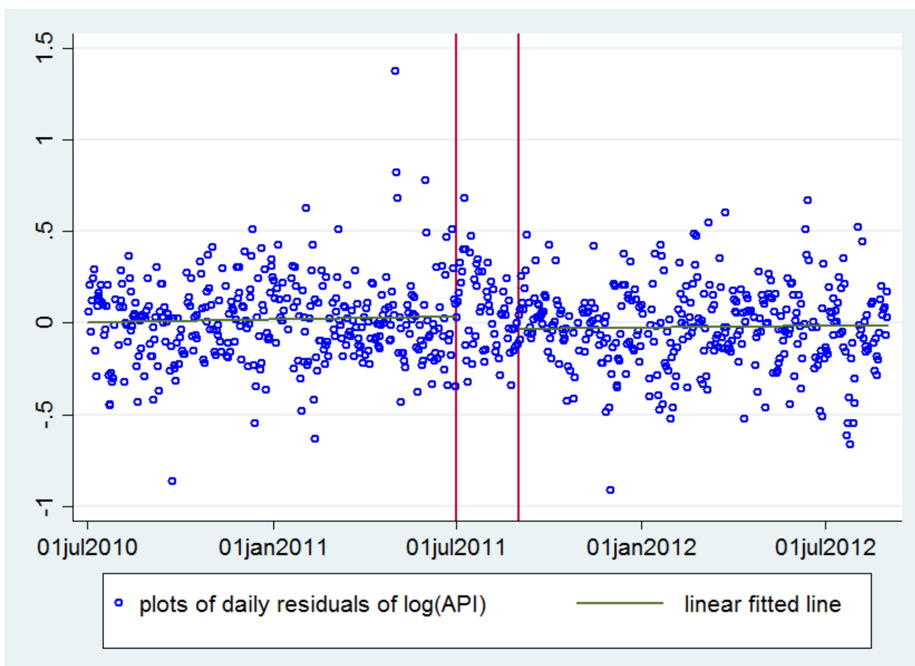


Figure 2.7: Daily Log API Residuals for the City of Hangzhou

2.4.4 Alternative Specification: Difference-in-Difference Method (DID)

The identifying assumption for the RD estimation is that, without the new fuel standards, air quality would not have discontinuously changed during the phase-in period. For example, the assumption would be violated if emissions from industrial sectors changed discontinuously due to some contemporaneous policy shift. To the best of my knowledge, I am not aware of any such policy introduced on a national scale at the same time as the fuel standards. However, as this assumption is fundamentally untestable, I complement my analysis with a difference-in-difference specification.

This approach takes advantage of the fact that cities upgrade their fuel quality at different paces. While the majority moved from China II to China III only when the new standard came into effect, four cities have implemented it earlier than that. The fact that those cities upgrade gasoline and diesel at the same time makes it impossible to separate their respective effects for these cities. Nonetheless, they can serve as a control group to form the counterfactual of how air quality would have evolved without changes in fuel standards during the national implementation period. If other national policies came into effect at the same time, it would have affected all the cities, thus leaving the differences unchanged. Therefore, we can attribute the changes in the differences to the new fuel standards.

One caveat regarding using this approach is that, as one may notice, the cities that have implemented CHINA III standards earlier are the most developed metropolises in China. As a result, they may experience forces that cause air pollution to increase or decrease over time relative to other cities. For example, economic activity may grow more quickly, which tends to be more polluting. Or these areas may also be more aware of environment issues and thus undertake more strict pollution control actions. Therefore, the parallel trend is less likely to hold for an extended period of time. To address this concern, I focus on a narrower 6-month window around the implementation dates for this analysis. Moreover, in some of the specifications, I allow city-specific linear or quadratic trends to distinguish the impacts of fuel regulations from individual trends driven by unobservables.

$$\ln(API_{ct}) = \alpha \cdot Treat_{ct} + \boldsymbol{\beta} \cdot \boldsymbol{W_{ct}} + \boldsymbol{\lambda} \cdot \boldsymbol{D_t} + \mu_c + \delta \cdot Trend_{ct} + \varepsilon_{ct} \tag{2.4}$$

The model is given by equation (2.4). $Treat_{ct}$ is a a variable indicating the implementation status of CHINA III gasoline or diesel. $Trend_{ct}$ captures city-specific linear or quadratic trends. W_{ct} and D_t include the same set of variables as in the RD setting. For inference, I allow the error terms to be correlated within the same city.

	Dependent Variable: ln(daily API)				
	(1)	(2)	(3)	(4)	(5)
CHINA III gasoline	0.035	-0.042	-0.049**	-0.046***	-0.047***
	(0.039)	(0.031)	(0.023)	(0.008)	(0.008)
City FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
Weather controls	N	Y	Y	Y	Y
DOW FEs	N	N	Y	Y	Y
Holiday	N	N	Y	Y	Y
Heating	N	N	Y	Y	Y
City-specific trends	N	N	N	Y	Y
City-specific quad trends	N	N	N	N	Y
N	10,618	10,098	10,098	10,098	10,098
R^2	0.304	0.518	0.522	0.551	0.571

Notes: Standard errors in parentheses are clustered by city. All effects are relative to the omitted baseline of CHINA II gasoline. The sample period is from November 1, 2009 to April 30, 2010.

*** Significant at the 1 percent level

** Significant at the 5 percent level

* Significant at the 10 percent level

Table 2.7: Difference-In-Difference Estimation Results: CHINA III Gasoline

The DID estimates are given in Table 2.7 and Table 2.8 for gasoline and diesel, respectively. The top row in each table reports the estimated effect of CHINA III standards. Moving from specification (1) to (5), I progressively add control variables which include weather, day-of-week variables, indicators for holidays and heating seasons, as well as city-specific time trends. Without these controls, neither of the effects is significant. However, after including city-specific time trends, the effect becomes statistically significant for CHINA III gasoline, while the effect of CHINA III diesel remains insignificant. The point estimates suggest that the introduction of CHINA III gasoline standard reduced API by 4.7%. Although the magnitudes of the estimates differ somewhat from the RD results, the overall pattern is very similar.

	Dependent Variable: ln(daily API)				
	(1)	(2)	(3)	(4)	(5)
CHINA III diesel	-0.052	-0.038	-0.038	-0.018	0.012
	(0.066)	(0.053)	(0.053)	(0.048)	(0.048)
City FEs	Y	Y	Y	Y	Y
Month FEs	Y	Y	Y	Y	Y
Weather controls	N	Y	Y	Y	Y
DOW FEs	N	N	Y	Y	Y
Holiday	N	N	Y	Y	Y
Heating	N	N	Y	Y	Y
City-specific trends	N	N	N	Y	Y
City-specific quad trends	N	N	N	N	Y
N	10,796	10,099	10,099	10,099	10,099
R^2	0.282	0.482	0.484	0.526	0.560

Notes: Standard errors in parentheses are clustered by city. All effects are relative to the omitted baseline of CHINA I diesel. The sample period is from May 1, 2011 to October 31, 2011.

*** Significant at the 1 percent level

** Significant at the 5 percent level

* Significant at the 10 percent level

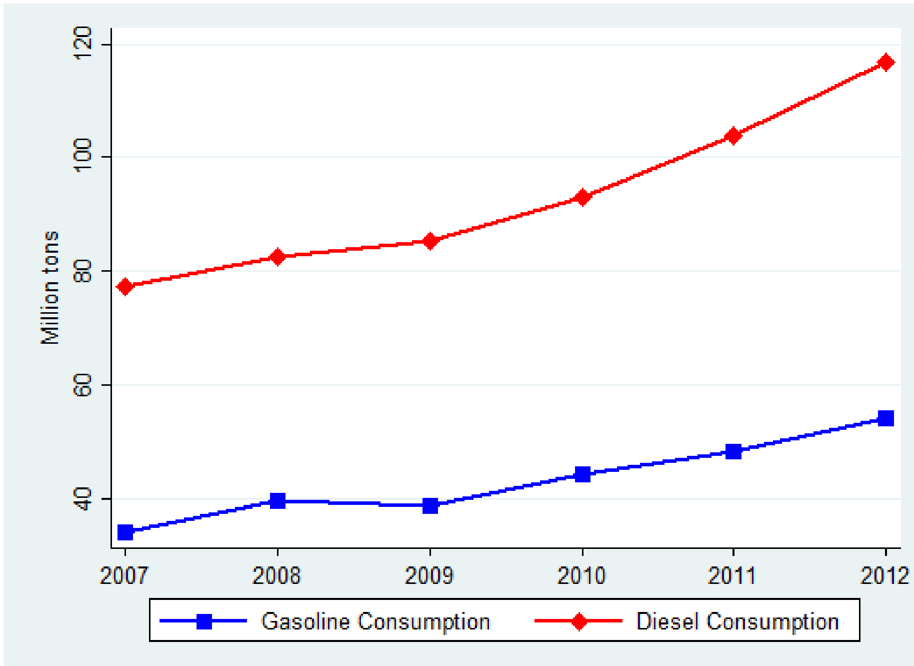
Table 2.8: Difference-In-Difference Estimation Results: CHINA III Diesel

2.5 Interpretation

2.5.1 Why Is CHINA III Diesel Standard Ineffective?

The finding that the effect of CHINA III diesel standard is insignificant and smaller than that of CHINA III gasoline standard seems perplexing, especially in view of the following facts. First, the reduction in the maximum sulfur level is 1650 ppm for diesel, relative to 350 ppm for gasoline. Second, particular matters are the major pollutant 70% of the time in China, and it has long been understood that diesel engines in general emit more NO_x and particular matters than gasoline engines.[14] Finally, more diesel is consumed than gasoline. Figure 2.8 shows the consumption quantities of gasoline and diesel in the transportation sector in China from 2007 to 2012. The amount of diesel consumed is roughly twice the amount of gasoline.

[14]Please refer to this website: http://www.isuzu.co.jp/world/technology/clean/diesel_gasoline01.html.



Data Source: National Bureau of Statistics of China

Figure 2.8: Gasoline and Diesel Consumption in the Transportation Sector: 2007-2012

Based on all the aforementioned facts, we should expect the diesel upgrade to have an effect at least as large as that of gasoline. The lack of an effect for CHINA III diesel seems to suggest that the new standard might not have been complied with as it should be. In this section, I consider two possible explanations for the different compliance responses from petroleum companies.[15]

2.5.1.1 Cost Difference in Desulfurization with No Pass-through

In China, the regulatory authorities of fuel quality are spread among several agencies under the state council. The structure is illustrated in Figure 2.9. Fuel quality standards are issued by the Standardization Administration of China (SAC) under the General Administration of Quality Supervision, Inspection and Quarantine (AQSIQ), while fuel prices are regulated by National Development and Reform Committee (NDRC)[16]. The decentralized structure

[15] In China, the petroleum industry is controlled by three state-owned vertically-integrated enterprises, namely, PetroChina, Sinopec and CNOOC.

[16] The Technical Committee under SAC in charge of drafting fuel quality standards is known as TC280, which consists of experts from the three major state-owned petroleum companies, the MEP and the auto industry. More than 90% of the members in TC280 are from petroleum companies, including the chairman of the committee. The secretariat of TC280 is based at Sinopec petrochemical research institute.

created coordination difficulty on fuel quality upgrades, as evidenced by the fact that the implementation of the new standards was not accompanied with either any fiscal help from the Ministry of Finance (MOF) or any adjustment in retail fuel prices by NDRC. Figure 2.4 confirms that there was no change in retail price caps when CHINA III gasoline and diesel standards were phased in.

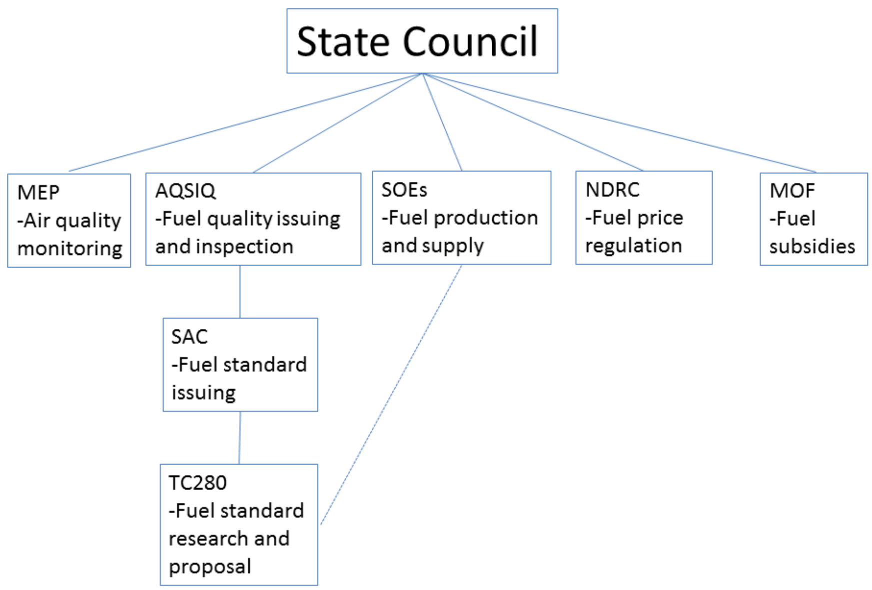


Figure 2.9: Structure of Government Bodies Related to Fuel Quality

	Gasoline	Diesel
Liu et al (2008)	0.9	1.3
Trans-Energy (2002)	1.1	2.1

Notes: All numbers are in US cents per gallon.

Table 2.9: Estimated Incremental Cost of Desulfurization

Consequently, refineries have to bear the entire cost of fuel upgrades. Table 2.9 lists two estimates from the previous literature on the incremental costs for Chinese refineries, when the sulfur level is reduced from 500 ppm to 150 ppm for gasoline and from 2000 ppm to 350 ppm for diesel. Both estimates suggest that it is more costly to upgrade diesel than gasoline. The cost difference amounts to 66.4 million dollars in 2011.[17] Without cost

[17] According to National Bureau of Statistics of China, in 2011, the total production quantities of gasoline and diesel are 79.2, 156.9 million tons respectively. Using Trans-Energy (2002)'s estimates, the annual cost difference is 156.9*0.021*28.56-79.2*0.011*31.75=66.4 million dollars, where 28.56 and 31.75 are used to convert tons to gallons for gasoline and diesel.

pass-through, upgrading diesel would have exacerbated the financial difficulty that many refineries were already facing.[18]

2.5.1.2 A Loophole in the Diesel Policy: Differential Standards for On-road and Off-road Diesel

While the lack of pass-through provides non-compliance motive for refineries, the presence of a much relaxed standard for off-road diesel facilitates evasion behavior for diesel producers. Off-road diesel (a.k.a "general diesel" in China) is used for trailers, locomotives with internal combustion engines, construction machinery, vessels, generator sets, 3-wheelers and low-speed trucks. Non-road diesel was subject to a different national standard which allows the maximum sulfur level to be as high as 2,000 ppm until July 1, 2013. In reality, on-road and off-road diesel are indistinguishable to the naked eyes. At gas stations, both of them are just labeled as diesel. Most consumers are not aware of the differences of the two, and even for inspectors, telling them apart would be impossible without lab tests.

The less stringent standard for off-road diesel creates a loophole to circumvent new regulations, as off-road diesel can be used for on-road vehicles illegally without being detected. Since there is no counterpart for gasoline, this could explain the different outcomes we observe for fuel upgrades. Such a practice has been reported by the news media.[19] Due to the illegal nature of this behavior and data limitation, empirical evidence on evasion activities is hard to obtain. However, an analogous situation has been documented and studied by Marion and Muehlegger (2008). In the US, diesel fuel used for on-road purposes is taxed, while other uses are untaxed, creating an incentive for firms and individuals to evade on-road diesel taxes by purchasing untaxed diesel fuel and then using it for on-road purposes. In response to that, the IRS and the EPA required fuel dye to be added to all diesel fuel not meant for on-highway use. This innovation substantially reduced the cost of enforcement by allowing regulators to randomly test trucks through a simple visual inspection. Consequently, they find that sales of diesel fuel rose by 26 percent following the regulatory change, while sales of heating oil, which is an untaxed perfect substitute, fell by a similar amount. Their estimates shed light on the magnitude of evasion when on-road diesel and off-road diesel are hardly distinguishable.

[18]In 2011, the two dominant players in the refining industry, Sinopec and PetroChina, reported 567 and 911 million dollar losses from the refining business, respectively.

[19]The following article is one example: http://finance.sina.com.cn/leadership/mroll/20131029/175517153428.shtml (Retrieved July 15, 2015).

2.5.2 Economic Significance

It is also interesting to examine how large the realized air quality benefits are in economic terms. To do so, one need to link the estimates in air quality improvement with well-established health effects, and then convert the health benefits to dollar terms using an appropriate value of a statistical life.

While the epidemiological literature has studied extensively about the relationship between air pollution and a range of health outcomes, for the purpose of meaningful comparison, I only focus on studies examining the impact of air pollution exposure on Chinese population. Since API is an aggregated pollution index that is not commonly used in the epidemiological literature, here I use the results on PM_{10} to calculate the benefits. Two meta-analyses which systematically review the studies on the effects of short-term exposure to air pollution conclude that a 10 $\mu g/m^3$ (10%) increase in PM_{10} was associated with a 0.31-0.32% increase in daily total mortality (Lai et al, 2013; Zhang et al, 2013). This translates to 23,665 lives saved per year given the effect of CHINA III gasoline.[20] I apply the value of a statistical life of $128.0 thousand suggested by Wang and He (2010) to calculate the benefits in economic terms. This leads to a total value of lives saved of around 3.0 billion annually, which dwarfs the upgrading cost of only 28.2 million dollars.[21]

My estimate is likely to represent a lower bound of the full benefit of the improved gasoline quality for several reasons. First, the fact that primary pollutants are reported only when API is above 50 implies that I cannot recover the value of PM_{10} when the fuel upgrade reduces the pollution level below the threshold. Therefore, the selective attrition may result in an underestimation of the actual reduction in air pollution. Second, my estimate is only based on one pollutant and its health consequence, whereas the fuel upgrade is also very likely to reduce the levels of other pollutants, such as CO, HC, and SO_2, as discussed in section 2.2. Due to the unavailability of pollutant-specific pollution data, I am not able to quantify the changes in those pollutants and their health effects which could be a significant portion of the potential benefits. Third, the improved air quality should have led to lower morbidity rates, the benefit of which is also not captured in the calculation. Finally, although the evident reduction in air pollution reflects that the new gasoline standard has

[20]The average annual morality rate in China between 2010 and 2012 is 7.13 per thousand people. The average population is 1.347 billion. This yields an estimated annual reduction in deaths of 0.077/0.1*0.0032*0.00713*1.347 billion=23,665.

[21]The upgrading cost is estimated by multiplying the average production quantity of gasoline from 2010 to 2012 with the incremental cost per gallon suggested by Trans-Energy (2002). The result is 0.011 dollars/gallon* 80.8 million tons*31.75 gallons/ton=28.2 million dollars.

taken effect in many cities, the lack of a direct measure of enforcement leaves the question of whether there was full compliance open. If full compliance was not met, then there is room for the potential benefits to be even higher than what I have found.

2.6 Conclusion

This paper examines the effectiveness of China's fuel sulfur regulation. In particular, I study the impacts of CHINA III fuel standards on air quality. Using both a time-series regression discontinuity design and a difference-in-difference method, I find that the introduction of CHINA III gasoline standard substantially reduces air pollution, while there is no evidence of improvement as a result of CHINA III diesel standard. The results are robust to alternative specifications and measures. The ineffectiveness of CHINA III diesel standard suggests it may not have been complied with. I discuss two potential explanations for this difference in compliance behavior. First, desulfurization is more costly for diesel than for gasoline. When cost pass-through is not permitted for the upgraded fuel, the cost difference is likely to motivate refineries to comply less for diesel. Second, a less stringent standard for off-road diesel creates a loophole to circumvent new regulations. The fact that off-road and on-road diesel are indistinguishable to the naked eyes makes it much easier to cheat on diesel than gasoline.

With the estimated pollution reduction effects, a "back-of envelope" calculation indicates that the health benefits amount to 3.0 billion per year, completely offsetting the upgrading cost. This finding is consistent with the cost-benefit analyses conducted in the US and Europe, which show that the magnitude of benefits is at least ten times larger than that of costs (EPA, 1999). The failure to enforce CHINA III diesel standard incur substantial social costs. The benefit estimate from the gasoline upgrade also serves as a lower bound for the foregone benefit of a cleaner diesel.

These results speak to the importance of carefulness in fuel policy design when enforcement is inadequate. Without effective enforcement, evasion is likely to prevail as long as on-road diesel and off-road diesel are still subject to different standards. Simple regulatory innovations, like requiring red dye to be added to off-road diesel, can help reduce supervision cost and prevent cheating. Moreover, providing economic incentives can be an effective way to overcome weak institutions and make moderately functioning governance structures work more effectively. If adequate price differentials were allowed to compensate firms for the upgrading costs, this would not only encourage firms to comply with the new standards, but also create political pressure to address the loophole problem, because the group of market participants who already upgraded would share an economic interest in seeing those standards being enforced.

CHAPTER 3

Pollution Liability Insurance and Corporate Environmental Compliance: A Case Study of Shenzhen

3.1 Introduction

For more than three decades, the Chinese economy has been growing rapidly with an average annual GDP increase rate of more than 9%. In the same period, China has also witnessed the deterioration of environment and an increasing number of environmental accidents. To control environmental risks and ensure adequate compensation for environmental damages, environmental pollution liability insurance is introduced as a market-oriented instrument of environmental governance. In case of environmental accidents, the insurance provides compensation for personal injury and property loss, and covers other related costs such as contamination clean-up charges.

In this paper, we examine if the insurance requirement improves firms' environmental compliance performance as measured by the number of their environmental violations. Theoretically, the insurance mandate could have two countervailing effects. On the one hand, the insurance protects firms from liability for accidents, which raises concerns for moral hazard. Evidence of moral hazard is well-documented in the health and automobile insurance literature. For example, Cohen and Dehejia (2004) show that traffic incidents increase when compulsory insurance is introduced. On the other hand, insurance companies set premiums with the goal of forcing firms to internalize the external costs. The premium is determined based on several characteristics including location, production scale, industry and evaluation of risk management capability. The latter further depends on past environmental compliance records (MEP & CIRC, 2013). Rubinstein and Yaari (1983) and Rogerson (1985) have shown that when premiums can depend on past records in a repeated game between insurer and insured and the insured pay their own premiums, incentives to reduce future premiums counteracts the ill effects of moral hazard. As a result, the net effect of the mandate is ambiguous from a theoretical standpoint. Therefore, this paper focuses on providing empirical evidence on the effect of this policy.

Given that the insurance mandate has been carried out on a national scale only recently, we perform a case study on one of the pioneers in promoting the policy—the city of Shenzhen. We focus on the electroplating and circuit board manufacturing firms which account for 75 percent of the firms that are mandated to purchase the insurance in Shenzhen. For comparison, we also include two non-required industries: the paper-product industry and the textile and dyeing industry, and an adjacent city that is not subject to the mandate – the city of Dongguan. The empirical strategy is a triple difference estimation. We compare (i) Shenzhen vs. Dongguan (first difference), (ii) before and after the introduction of the liability insurance (second difference), and (iii) industries covered by this policy vs. not covered (third difference).

We use a novel dataset on corporate environmental performance that has not been exploited in the literature. Firms' compliance performance is measured by the number of environmental violations they commit in each year. Our empirical results suggest that the insurance requirement leads to a 73-97 percent reduction in the expected number of violations. The results are robust to a number of specification checks, including different count models and OLS estimation. Furthermore, the results are stable across different sample selection criteria and the data exhibit similar trends across cities prior to treatment, alleviating concerns about selection bias.

Our study is related to several empirical studies on environmental liability and accidents. Alberini and Austin (2002) exploits cross-state variation in mini-superfund programs to estimate the effect of the programs on uncontrolled toxic releases in the United States, and find that strict liability reduces the frequency and severity of pollution releases. Yin et al. (2011) shows that mandating environmental liability insurance for underground storage tanks can address the inefficiency due to small firms declaring bankruptcy. For the same reason, bonding requirements for oil companies decrease environmental accidents (Boomhower, 2016). This paper adds to our understanding of the effect of liability policies in the Chinese context. This paper also contributes to a growing body of literature showing that mandatory insurance policies can be used as a beneficial form of environmental regulation (Katzman, 1988; Kolstad et al., 1990; Farber, 1991; Zweifel and Tyran, 1993;

Ben-Shahar and Logue, 2012).

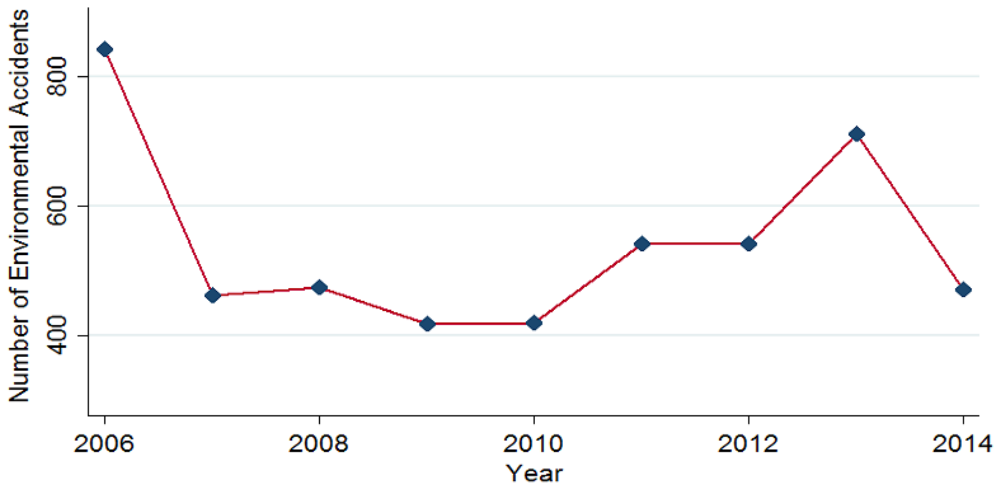
The rest of the chapter is organized as follows: Section 3.2 provides background information regarding environmental accidents and the pollution liability insurance in China. Section 3.3 presents a simple model of firms' optimal safety effort with and without the insurance. Section 3.4 describes the data. Section 3.5 discusses the results, and section 3.6 concludes.

3.2 Policy Background

3.2.1 Environmental Accidents in China

China's remarkable economic and industrial development is accompanied with worsening environmental quality and a growing number of environmental accidents. Figure 3.1 shows the number of environmental accidents that occurred annually from 2006 to 2014 (China Environment Yearbook, 2007-2015). On average, there were roughly 540 accidents happening each year, many of which were catastrophic, resulting in water pollution, farmland contamination, poisoning and even death. Table 3.1 presents several major environmental accidents from 2000 to 2010. These accidents have led to productivity loss in farmland and fisheries, and exposed thousands of Chinese citizens to health risks. The direct financial loss was estimated to be more than tens of millions of dollars (China Environment Yearbook, 2010). The cost to clean up the damages and restore the ecosystem was even beyond estimation.

These accidents have also given rise to social unrest. For victims, getting timely and fair compensation for health and property damages is challenging. This is in part due to the inadequacies of China's legal system when it comes to environmental issues. Litigation is often very lengthy, and evidentiary burdens are sometimes unreasonable. Moreover, establishing a causal effect and assessing the damages requires special expertise. When it comes to health issues, compensation is often inadequate, owing to a lack of precedent for quantifying damages. For example, when a chemical company in Hunan was found to be responsible for blood lead levels in 13 children that exceeded China's national standard, families of 11 of the children were denied compensation on the grounds that their condition did not require medical procedures or drugs and thus the court had no procedure for calculating damages due to lead in the blood. The two children with the highest blood levels were able to obtain only 10,000 RMB, or approximately $1,600, in compensation.[1] The outrage of the public sometimes results in mass demonstration and protests. In those cases, the government has to step up to handle the damages and reassure the victims before it turns into a public crisis.[2]



Source: China Environment Yearbook, 2007-2015

Figure 3.1: Number of Environmental Accidents: 2006-2014

Environmental Accident	Date	Polluting Firms	Province	Consequence
Duyun dam collapse	9/11/2002	Lead and Zinc Mining Firms	Guizhou	Water and land contamination
Aniline explosion and leakage	11/13/2005	Jilin Petrochemical Company	Jilin	6 killed; 70 injured; River contamination
Lead poisoning	8/2/2009	Dongling Group	Shaanxi	851 children poisoned
Cadmium pollution	8/6/2009	Xianghe Chemical Plant	Hunan	26 killed; Many poisoned; Farmland contamination
Copper acid water leakage	7/3/2010	Zijin Mining Group	Fujian	River contamination; Financial loss of $4.6 million from fisheries
Xingang oil spill	7/16/2010	Dalian New Port	Liaoning	Marine contamination

Source: Various news reports

Table 3.1: Major Environmental Accidents in China: 2000-2010

[1] See the news report at http://view.news.qq.com/original/intouchtoday/n3448.html.

[2] News coverage on protests can be easily found online, such as this one: http://www.europe-solidaire.org/spip.php?article15399.

3.2.2 Environmental Pollution Liability Insurance

To deal with growing public concern about the environment and ensure fair compensation for environmental damages, the Chinese government introduced environmental pollution liability insurance as a new economic instrument. Environmental liability policies originated in industrialized countries in the 1960s. One prominent example of such regulation is the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), also known as the Superfund, passed by the US Congress in 1980. Under CERCLA, the Environmental Protection Agency (EPA) can identify parties responsible for hazardous substances releases to the environment and either compel them to clean up the sites, or undertake the cleanup on its own using the Superfund (a trust fund). The EPA may recover the costs by working with the U.S. Department of Justice to pursue the responsible party through the court system. Approximately 70 percent of Superfund cleanup activities have been paid for by parties responsible for the cleanup of contamination.[3]

In China, the development of the environmental pollution liability insurance involves three stages (Feng et al., 2014). The insurance was first introduced to a few northeastern cities in 1991 (Congjun and BinChik, 2012). However, the enrollment was low, as only 15 firms had purchased coverage between 1991 and 1994 (Born and Chen, 2013). The second phase took place in 2007 when "The Guidelines on Environmental Pollution Liability Insurance" (The Guidelines) was issued jointly by the Ministry of Environmental Protection (MEP) and the China Insurance Regulatory Commission (CIRC). The Guidelines encouraged local governments to experiment with promoting the insurance. Four provinces and four cities were chosen as pilots by the end of 2008.[4] These places represent different geographic regions of China. By 2012, 14 cities have launched trial applications of the policy. The third phase started in February 2013 when the central government promulgated "The Guiding Opinions on Pilot Scheme for Compulsory Environmental Pollution Liability Insurance" (The Opinions). The Opinions mandated the insurance to be purchased by all firms dealing with heavy metals. The insurance was also recommended for certain other industries (see Table 3.2). According to The Opinions, three types of expenses are covered by the insurance in case of accidental pollution: (1) third-party liability (personal injury, death and property loss), (2) necessary and reasonable expenses incurred by the insured to save a third party's life (including expenses for medical treatment) or to prevent or mitigate the loss of property of any third party, and (3) necessary and reasonable clean-up expenses incurred by the insured in order to control the extent of pollution or to remediate contami-

[3]U.S. EPA, Factsheet, "Superfund Trust Fund and Taxes: Setting the Record Straight," October 7, 2003.
[4]These are Jiangsu, Hubei, Hunan, Henan, Chongqing, Shenzhen, Ningbo, and Shenyang.

nated waters and land in accordance with environmental legislation.

Required	Miners and processers of heavy non-ferrous metal ore
	Heavy non-ferrous metal smelting
	Lead battery manufacturing
	Leather and leather product manufacturing
	Chemical raw material and chemical product manufacturing
Recommended	Petrochemical
	Producers, warehousers, users and transporters of dangerous chemicals
	Hazardous waste treatment plants
	Industries with dioxin emission

Table 3.2: Industries Required or Recommended to Purchase the Pollution Liability Insurance

Guided by The Opinions, each province is responsible for working out its own implementation plan based on local conditions. Therefore, there is considerable variation across regions in terms of roll-out pace, enforcement strength and industry coverage. As of 2017, about two-thirds of the provinces have implemented the policy in some capacity. To motivate firms to purchase the insurance, compliant firms are generally given access to special environmental protection funds and priority in bank lending, while non-compliance triggers sanctions, such as negative environmental impact assessments, suspension of access to special environmental protection funds and credit downgrades.

In this paper, we study whether the mandatory pollution insurance policy improves firms' environmental compliance. Although the effect on environmental accidents is interesting, we do not have enough observations on accidents given that they are rare in nature.[5] Instead, we focus on how the mandatory pollution insurance affects the relatively "small" violations of environmental regulations. On one hand, having insurance can lead to moral hazard and cause firms to take less care in their production process. This will result in more violations. On the other hand, two channels exist to incentivize polluters to improve their operation and increase compliance with environmental legislation. The first channel is through the environmental risk assessment conducted by a third party as required by the government. Third party experts evaluate firms' characteristics such as production scale, location, and operational procedures and determine their riskiness. More risky firms will be required to purchase insurance with higher coverage at a higher expense.[6] They can also

[5]For example, in Shenzhen, the first insurance payout happened in 2015 – eight years after the initial implementation.

[6] The coverage requirement can range from 150k to 1.5 million dollars.

provide suggestions on how to reduce their risks. For example, if loopholes in pollution treatment are identified during the assessment, they can advise firms on how to fix them. The second channel lies in how the premiums are structured. Insurance companies take into account firms' past environmental compliance records when setting their premiums. Firms with no violation records will be offered discounts. This provides another channel to incentivize firms to comply with environmental regulations.

3.2.3 A Case Study of Shenzhen

We perform a case study of Shenzhen to examine the effect of the mandatory pollution insurance policy on firms' environmental compliance. The city of Shenzhen is the only pilot city in South China, and one of the first to implement the insurance policy on a large scale. It is located in the relatively developed Pearl River Delta, bordering Hong Kong to the south.

Shenzhen is selected as a pilot largely due to its status as an experimental field for both economic reform and environmental protection. Shenzhen was singled out to be the first special economic zone of China in 1980, and has proven to be one of the most robust and fastest growing cities in the country since then. Shenzhen is also regarded as a "green" city. It was named the nation's first model city for environmental protection in 1997, and thereafter won other awards such as "model city for protection of ozone layer" and "national greenery model city." It has been a test field for a number of environmental policies. For example, in 2013, Shenzhen became one of seven cities to pilot China's regional carbon emission trading system.

Shenzhen is a suitable case for the empirical analysis for two reasons. First, being a pilot city, Shenzhen has a large number of insured firms and a relatively long history of implementing the policy. The mandatory insurance requirement was first introduced in 2008 for firms producing hazardous waste, and then extended to other industries such as hazardous chemical and lead battery producers, sewage and garbage disposal plants, and electroplating and circuit board factories in 2012. The coverage reflects Shenzhen's industrial structure and includes firms that pose the highest environmental risks to the public. As of 2015, a total of 747 firms were subject to the insurance mandate in Shenzhen. Second, Shenzhen is one of the few regions that publicly list the names of required and insured firms and continuously update the lists from year to year. This allows us to keep track of each firm's status (e.g. whether a firm was required to purchase insurance and whether it actually purchased it in each year) and associate it with the environmental violation records of that firm. For the analysis, we focus on the electroplating and circuit board manufacturing industry (hereafter referred to as ECB). ECB companies account for about 75 percent of all the required firms. Aside from having a large number of firms, they also produce relatively homogeneous products. [7]

For comparison, we use the city of Dongguan as a control group. Dongguan is adjacent to Shenzhen and has a similar industrial structure. Figure 3.2 shows the location of the two cities. Dongguan did not impose any insurance requirement until 2016. We exploit this difference in implementation speed to examine the effect of the insurance mandate. We also include the paper-product industry and the textile & dyeing industry as controls.[8] These two industries are also regular polluters but are not required to buy the insurance due to their relatively low risk of causing environmental disasters.

3.3 Model

This section presents a simple conceptual model of firms' optimal safety effort and compliance behavior with and without the mandatory pollution liability insurance. Suppose a competitive hazardous industry is comprised of risk-neutral homogeneous firms. Firms may exert costly safety effort e to comply with environmental regulations. The level of safety effort affects both the number of environmental accidents, and the number of relatively "small" environmental violations which is the variable of interest in the empirical section.[9] The number of accidents and the number of violations are both random variables with means $\mu_a(e)$ and $\mu_v(e)$, respectively. We assume $\mu_a'(e) < 0$, $\mu_a''(e) > 0$, $\mu_v'(e) < 0$, $\mu_v''(e) > 0$. In the case of an environmental accident, the social damage is C_s, while firms

[7] By contrast, although hazardous chemical producers are also required to purchase the insurance, depending on which chemicals they produce, the production process can be very different, making direct comparison across firms less viable.

[8] We compile the list of firms in those industries by selecting firms which pay pollution discharge fees (indicating they are polluters) and have certain relevant characters in their names. For example, we will pick out polluting firms that contain "paper" in their names and put them under the "paper-product industry" category. This is not perfect as not all paper-product companies have "paper" in their names. But if whether containing "paper" or not in the name is uncorrelated with the other characteristics of the firm, this process gives us a random sample of firms in this industry.

[9] We assume that accidents and violations are always detected. For simplicity, we also assume fines are equal to damages for violations. The results still hold qualitatively if we assume fines are instead a fraction of social damages.

Figure 3.2: Location Map of Shenzhen and Dongguan

pay C_f, $0 < C_f < C_s$, due to reasons described in Section 3.2.1. The difference between C_f and C_s is either borne by the victims or paid by the government. For simplicity, we assume both C_f and C_s are constant. That is, the safety effort affects only the likelihood of an accident but not the severity of an accident. In the case of an environmental violation, firms have to pay fines F which are also assumed to be constant. The cost of effort is $T(e)$ with $T'(e) > 0$ and $T''(e) > 0$.

First, consider a case where there is no liability insurance requirement. A firm chooses e to minimize its overall cost, which consists of effort cost and environmental cost:

$$T(e) + \mu_a(e)C_f + \mu_v(e)F$$

The optimal safety effort e^* satisfies the first-order condition $T'(e) + \mu_a'(e)C_f + \mu_v'(e)F = 0$. On the other hand, the socially optimal effort level e^s minimizes $T(e) + \mu_a(e)C_s + \mu_v(e)F$ and satisfies $T'(e) + \mu_a'(e)C_s + \mu_v'(e)F = 0$. Intuitively, since firms have to pay only a portion of the actual damages related to environmental accidents, they will exert insufficient effort to comply with environmental regulations. A formal proof of this statement is provided as follows.

Proposition 1 *With no pollution liability insurance, the optimal safety effort e^* is less than the socially optimal effort e^s. Hence, the expected number of environmental violations is higher than the socially optimal level.*

Proof: Given $0 < C_f < C_s$, the first order conditions imply that

$$-\frac{T'(e^*) + \mu_v'(e^*)F}{\mu_a'(e^*)} < -\frac{T'(e^s) + \mu_v'(e^s)F}{\mu_a'(e^s)}.$$

It is easy to show that $\frac{T'(e)+\mu_v'(e)F}{\mu_a'(e)}$ is a decreasing function of e. Hence, we have $e^* < e^s$. Since $\mu_v(e)$ is a decreasing function of e, it follows that $\mu_v(e^*) > \mu_v(e^s)$. *Q.E.D.*

Now let us consider how the mandatory pollution liability insurance changes firms' incentives. The insurance protects firms from liability for environmental accidents but not for environmental violations. Firms purchase this insurance from insurers in a competitive market. Each firm pays a premium R and gets full coverage in case of environmental accidents. If insurers are naive in the sense that they ignore moral hazard, they will charge a fair premium R equal to the expected value of the social damages $\mu_a(e^*)C_s$. If insurers are sophisticated, they will anticipate moral hazard and set up a premium structure that incentivizes safety effort. The premium can not depend on firms' safety effort directly, since effort may not be fully observable. However, insurers can adjust rates based on firms' accident and regulatory compliance history.

Proposition 2 *With the pollution liability insurance, when insurers set fixed premiums, firms exert less safety effort than e^*. As a result, the expected number of environmental violations exceeds $\mu_v(e^*)$. However, a properly designed premium structure can incentivize firms to exert the socially optimal effort level.*

Proof: If premiums are fixed, a firm chooses e to minimize $T(e) + R + \mu_v(e)F$. Note that R is not a function of e. Therefore, the optimal effort level e^n satisfies $T'(e^n) + \mu_v'(e^n)F = 0$. Since $T'(e^*) + \mu_a'(e^*)C_f + \mu_v'(e^*)F = 0$ and $\mu_a'(e^*) < 0$, we have

$$T'(e^*) + \mu_v'(e^*)F > T'(e^n) + \mu_v'(e^n)F.$$

Given that $T'(e) + \mu_v'(e)F$ is an increasing function of e, $e^n < e^*$. As a result, the expected number of violations increases, i.e., $\mu_v(e^n) > \mu_v(e^*)$.

However, if R is a function of the number of violations v, a firm chooses e to minimize its overall cost including the expected premium, $T(e) + E(R(v(e)) + \mu_v(e)F$. For certain R

functions, the optimal effort may be higher than e^*. In particular, a R function that satisfies $E(R(v(e))) = \mu_a(e)C_s$ can induce the socially optimal effort level. *Q.E.D.*

In sum, whether the number of environmental violations increases or decreases as a result of the mandatory pollution liability insurance depends on whether the insurance provides sufficient incentive to combat moral hazard. In the next sections, we empirically examine the effect using data from Shenzhen and Dongguan.

3.4 Data

To measure firms' environmental performance, we use a database from the Institute of Public & Environmental Affairs (IPE).[10] IPE is a non-profit environmental research organization registered and based in Beijing, China. Since its establishment in June 2006, IPE has dedicated itself to collecting, collating and analyzing government and corporate environmental information. The IPE database provides a wealth of information regarding environmental violations committed by companies and factories. The records are drawn from various sources, including reports from news articles and local environmental protection agencies. Each record identifies the name of the polluter, describes the violation, and documents the date, the supervision agency and the record source.

We use the number of violations a firm commits in a year to measure its environmental performance. A violation can be any misconduct from exceeding pollution limits to operating pollution treatment devices inappropriately or failing to obtain environmental permits for new projects. Some violations are detected through onsite inspection, while others are found using automatic monitoring devices. We remove two types of records. First, we remove entries related to firms' annual environmental credit rating grades, as they reflect firms' overall environmental performance in a given year and hence double count any violations firms have already committed. Second, we remove any records that are the results of special enforcement actions. Over time, local environmental protection agencies may take special enforcement actions that temporarily intensify the inspection effort, and thus affect the number of detected violations. To rule out this effect, we go over the cases very carefully and remove all the records that are the results of such actions. This leaves us with 1,933 violation records for 1,064 firms in the sample from 2006 to 2014.

[10] http://www.ipe.org.cn/

City	Industry	Number of Firms	Average Number of Violations
Shenzhen	ECB industry	557	0.10
	paper-product industry	14	0.06
	textile & dyeing industry	29	0.07
Dongguan	ECB industry	165	0.13
	paper-product industry	128	0.09
	textile & dyeing industry	171	0.06

Notes: This table reports the number of firms and the average number of violations per firm per year for each industry in each city based on data from 2006 to 2014.

Table 3.3: Summary Statistics of the Data

Table 3.3 presents summary statistics for the three industries in Shenzhen and Dongguan. The numbers vary across cities and industries. On average, for every 100 ECB firms, 10 violations are committed per year in Shenzhen, while 13 violations are committed per year in Dongguan. To motivate the regression analysis, we plot how the average number of violations changes over time for different groups as shown in Figure 3.3. This figure allows us to examine pre-treatment trends in the number of violations. In general, the movements in Shenzhen and Dongguan track each other closely before 2012. The number of violations was low before 2012 and increased substantially after 2012. Due to the lack of data, the cause of this sudden increase is not entirely clear, but is likely to be a result of changes in several factors. The most likely factor contributing to the increase is the delegation of supervision power to lower levels. Before 2013, only the environmental supervision branch directly under the municipal environmental protection agency has the punitive power, but afterwards this power was delegated to all protection bureaus at districts and towns. This change increased the number of supervisory personnel, and thus could lead to a sudden increase in the number of violation records.[11]Other factors that may have contributed to the increase include greater scrutiny on polluters and the increasing usage of automatic monitoring devices.[12] Because of these possible changes, a direct before-after comparison will be problematic and therefore we focus on the difference in numbers across cities. For ECB firms, the gap became larger after 2012, which suggests that the policy might have an impact. However, the change in the gap could have been driven by an idiosyncratic shock to one city but not the other around the same time as the policy was implemented. Indeed,

[11]A principal at the policy and regulation division of Dongguan's environmental protection agency confirmed this in an interview with Dongguan Daily. See the news at `http://news.sun0769.com/dg/headnews/201311/t20131106_2985615.shtml`.

[12]In 2006, none of the violations was detected by automatic monitoring devices, while in 2014, about 10% were. Automatic monitoring devices make detection of violations easier and hence can explain part of the increase in the number of violations.

we do see a wider gap for the industries in which firms were not required to purchase insurance after 2012 as well. This motivates us to look at a third difference – the difference between required and non-required industries. This will take care of any city-level shock that affects the number of violations in all three industries.

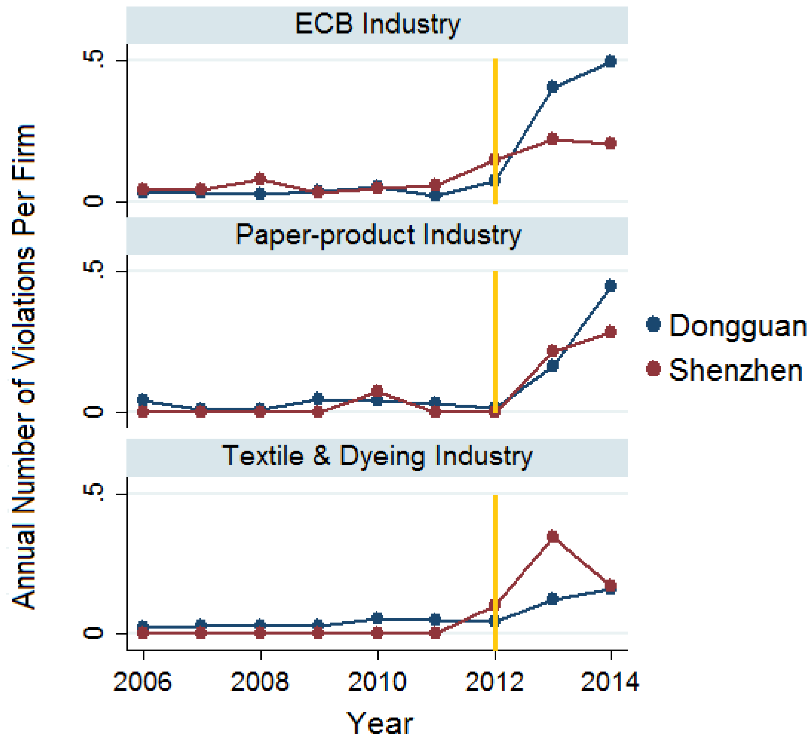


Figure 3.3: Average Number of Violations by Year, Industry and City

3.5 Empirical Evidence

The empirical strategy is conceptually a triple difference estimation. We want to compare (i) Shenzhen vs. Dongguan (first difference), (ii) before and after the introduction of the liability insurance (second difference), and (iii) industries covered vs. not covered by this policy (third difference). Specifically, we estimate the following equation:

$$\begin{aligned} Vio_{it} =& \alpha Shenzhen_i + \beta ECB_i + \mathbf{Year_t} + \theta Shenzhen_i * ECB_i + ECB_i * \mathbf{Year_t} \\ &+ Shenzhen_i * \mathbf{Year_t} + \gamma Shenzhen_i * ECB_i * Post_t + \epsilon_{it} \end{aligned}$$

Vio_{it} is the number of violations firm i committed in Year t. $Shenzhen_i$ is a dummy indicating if firm i is in Shenzhen. ECB_i is a dummy indicating if firm i is in the ECB industry. $Year_t$ are dummies for year fixed effects, and $Post$ is equal to one for years after 2012 and zero otherwise. γ is the measure of the treatment effect we are interested in. We present results from three models: OLS, a poisson model and a negative binomial model. The negative binomial model is most preferred given its capability of dealing with overdispersed count data.[13] For all specifications, standard errors are clustered at the industry-city level to allow for arbitrary correlation across time and across firms within the same industry and city.

	Full Sample			Excluding 2012	Excluding Entries & Exits	Excluding non-compliant Firms
Model	OLS	Poisson	Neg bin	Neg bin	Neg bin	Neg bin
Treatment effect	-0.30***	-3.66***	-3.64***	-3.83***	-3.67***	-3.63***
	(0.04)	(1.15)	(1.15)	(1.13)	(1.15)	(1.15)
Change in expected value (%)	-73	-97	-97	-98	-97	-97
Number of observations	9,576	9,576	9,576	8,512	8,244	7,749

Notes: Treatment is considered as "being required to purchase the pollution liability insurance". Treatment effect corresponds to the estimate of coefficient γ in the model.

*** Significant at the 1 percent level

Table 3.4: Effect of the Mandatory Pollution Liability Insurance on Environmental Violations

Results are shown in Table 3.4. To compare the magnitude in the OLS specification with the magnitude in the count specifications, we also present the percentage change in the expected number of violations attributable to the policy for all regressions.[14] With the full sample, for all three specifications, the treatment effects are negative and statistically significant at the 1% confidence level. In the OLS specification, the insurance requirement leads to a 73 percent reduction in the expected number of violations, while in the count models, the estimated change is even larger, a drop of 97 percent.

[13]The Poisson process assumes equality of the mean and variance, whereas in empirical settings the variance is often larger than the mean. This overdispersion leads to faulty inference. In our sample, the overdispersion parameter is estimated to be 1.38, significantly different from 0. Hence, we can reject the poisson model in favor of the negative binomial model. Nevertheless, we include the poisson regression results for robustness check.

[14]For the count specifications, the percentage change in the expected number of counts is equal to $exp(\beta) - 1$. For the OLS specification, the percentage change in the expected number of counts is equal to $\frac{\beta}{E(Y_i|Shenzhen=1,ECB=1,Post=0,Year\geq 2012)}$.

The rest of the columns in Table 3.4 present results from three robustness checks. The first robustness check excludes observations in year 2012. The mandatory insurance requirement for ECB firms was introduced on May 30, 2012. Therefore, those firms had no obligation to purchase the insurance for the first 5 months of 2012. Additionally, it may take some time for the policy to take effect. Therefore, we exclude all observations in year 2012. The second robustness check only includes the ECB firms in Shenzhen that are consistently required to purchase the insurance from 2012 to 2014. The list of required firms was adjusted every year to account for entries, shut-downs, or exits. Of the 557 required firms in Shenzhen, 409 firms were on the required list every year from 2012 to 2014. So for the second robustness check, we only include those firms. Finally, we exclude non-compliant firms. While we consider the treatment as "being required to purchase the insurance", in reality not every firm is in compliance with this policy and the effect can differ based on whether they have actually purchased the insurance or not. About two fifths of the required firms have not purchased the insurance as of 2014. In the last column, we present the results where only the compliant firms are included. Overall, the results suggest that the estimates are very robust to these different sample selection criteria.

3.6 Conclusion

This paper provides empirical evidence on the effect of mandatory pollution liability insurance on firms' environmental compliance performance. The total drop in violation incidents is estimated to be 73-97 percent and statistically significant at the 1 percent level. As the insurance policy continues to roll out on a national scale, this paper provides a timely evaluation of its effectiveness.

Several caveats apply. First, we estimate the treatment effect on the treated. The results may not directly apply to other regions and industries. In particular, as the model suggests, the effect depends on the premium structure. Although I find the insurance reduces violations in Shenzhen, to the extent that other regions may adopt a different premium formula, firms' compliance behavior may be different in those places. Therefore, a cross-region or cross-industry comparison of the effectiveness of this policy will be a direction for future research. Second, for the analysis, we assume away any shock that affects only one industry in only one of the cities. We consider such a case to be unlikely, given that the cities are next to each other and firms in those cities serve a much larger national and even global market. However, this is a possibility that we cannot completely rule out. In particular, the number of violations is subject to enforcement strength. Although we have done our best to tease out special inspections, we cannot observe the enforcement effort directly. If enforcement effort changes in different ways for different industries, this may invalidate our approach.

APPENDIX A

ERCOT Market Operation

This section provides details on the market processes under the bilateral trading market and the centralized auction market.

A.1 Scheduling and Dispatch Under the Bilateral Trading Market

Before the redesign, ERCOT was a bilateral trading market. The operation of the market consists of two major phases.[1]

1. Day-ahead scheduling process
   Load serving entities and generation resources negotiate privately with each other to buy and sell energy. The resulting bilateral contracts specify the transfer of electricity at negotiated terms such as duration, price, and time of delivery. In the day-ahead period, market participants are required to submit their "balanced schedules" to the ERCOT ISO through Qualified Scheduling Entities (QSEs) which are qualified by ERCOT to submit schedules for a portfolio of generators and power purchasers. These schedules specify the origins and destinations of power flows by congestion zone for each 15-minute settlement interval.[2] The scheduled resource production should not deviate from the forecasted demand beyond an established range. ERCOT analyzes the day-ahead schedules and notifies the QSEs of anticipated inter-zonal congestion. Market participants are allowed to adjust their schedules to relieve the forecasted congestion. Once the schedules are accepted by ERCOT, the

[1] There is also an adjustment period between the day-ahead period and the operating period.

[2] ERCOT divides its territory into 4 congestion zones. A congestion zone is a group of buses that have similar shift factors on commercially significant constraints. Dividing the entire grid into several congestion zones simplifies the modeling of the network.

generators are "physically" committed to produce the scheduled quantity unless being instructed to increase or decrease their production in the balancing market. Any uninstructed deviation exceeding 1.5% or 5 MWh of the QSE's schedule results in a penalty payment (Sioshansi and Hurlbut, 2010). 95% of the overall generation is scheduled through this process.

2. Real-time balancing market
   During the day-ahead scheduling process, generation resources also submit balancing energy bids for adjusting their generation relative to their scheduled quantities. In real-time, ERCOT manages energy imbalance and transmission congestion between zones by intersecting the bidding functions separately for each zone. For intra-zonal congestion, ERCOT deploys resources based on the generic fuel cost factors and shift factors to resolve local transmission constraints.

A.2 Scheduling and Dispatch Under the Centralized Auction Market

Under the centralized market design, market participants put their generation resources at the disposal of ERCOT. These resources are centrally dispatched to minimize generation costs. The operation of the centralized auction market also consists of two phases.

1. Day-ahead operation
   In the day-ahead period, market participants submit offers to sell energy for each hour of the operating day. The supply offer may contain three parts: the startup offer, the minimum-energy offer and the energy offer curve. These offers are used in the day-ahead energy market. Participation in the day-ahead energy market is voluntary and does not physically commit a resource to come on-line. In 2011, day-ahead purchases account for approximately 40 percent of the real-time load (Potomac Economics, 2012). After the completion of the day-ahead energy market, ERCOT executes a reliability unit commitment process to ensure that it has enough capacity committed to serve the forecasted load for the operating day.

2. Real-time operation
   While bilateral trades and the day-ahead energy market transfer *financial* responsibility among QSEs, the Security Constrained Economic Dispatch (SCED) program actually dispatches the resources in the real time. ERCOT utilizes a network operation model which represents the system with critical information on characteristics,

ratings, and operational limits of all elements of the transmission grid. On-line resources are dispatched in economic order according to their submitted energy offer curves. The execution of SCED results in locational marginal prices at approximately 4,000 nodes.[3]

[3] Hence, the centralized market is also known as the "nodal market."

APPENDIX B

Proof of Firm X's Optimal Strategy

Denote the quantities Firm X supplies at node B and C as Q_B and Q_C. The supply coming from node A is $300 + Q_B$. Hence, the residual demand for Firm X at node C is $Q_C = 600 - (300 + Q_B) - Q_B - 100(P_C - 8) = 1100 - 2Q_B - 100P_C$. Equivalently, $P_C = 11 - 0.02Q_B - 0.01Q_C$. To obtain P_B, note that if we increase production at both A and B by 1 MW each, production at C can be reduced by 2 MW to meet the same level of demand at C. The resulting prices, therefore, satisfy the relationship P_A+P_B=$2P_C$. Firm X's problem is:

$$\max_{Q_B \geq 0, Q_C \geq 0} (11 - 0.02Q_B - 0.01Q_C - 7.2) * Q_C + [2 * (11 - 0.02Q_B - 0.01Q_C) - 5 - 9] * Q_B$$

The kuhn-Tucker conditions with respect to Q_B and Q_C are

$$\frac{\partial}{\partial Q_B} = 8 - 0.08Q_B - 0.04Q_C \leq 0$$
$$\frac{\partial}{\partial Q_C} = 3.8 - 0.04Q_B - 0.02Q_C \leq 0$$

Obviously, the equalities will not hold for both conditions. We must have

$$\frac{\partial}{\partial Q_B} = 0$$
$$\frac{\partial}{\partial Q_C} < 0$$

Hence, the profit-maximizing quantities are $Q_B^* = 100$, $Q_C^* = 0$.

APPENDIX C

Data Appendix

C.1 Coal Price

The majority of a power plant's coal is purchased through long-term contracts. Therefore, I use monthly plant-level coal receipt cost data from EIA-923 forms as the relevant coal prices. Some previous studies have used spot market coal prices to approximate the opportunity costs for coal plants (Mansur (2008), Mansur and White (2012)). However, spot market prices are not appropriate proxies for opportunity costs for two reasons. First, there is evidence that the pass-through from spot market price to contract price for coal is fairly long and incomplete. Chu et al (2015) find that a 1% change in the coal spot price leads to only an approximately 0.11% change in the contract prices received by power plants even after 12 months. Second, power plants consistently pay a sizable premium for contract coal over spot coal, which suggests that there are industrial or institutional barriers to taking advantage of the cheaper spot coal. Joskow (1987) and Jha (2014) attribute this phenomenon to transaction-cost economics and regulatory-induced risk aversion, respectively.

The fuel receipt cost data are publicly available for regulated plants.[1] There are 16 coal plants in ERCOT, 6 of which are regulated. For deregulated plants, I approximate the coal prices in the following way. Power plants in Texas purchase two types of coal: lignite from Texas and sub-bituminous coal from the powder river basin in Wyoming. Only 2 regulated plants purchase lignite. Since lignite is produced within Texas, I assume that the transportation costs are relatively small while the content of the coal matters more for the price. Hence, I use the coal prices paid by plant Pirkey to approximate the prices for deregulated plants, since the characteristics of coal purchased by Pirkey are close to the average lignite being purchased. However, for sub-bituminous coal, the transportation cost is likely to be important. Therefore, I match every deregulated plant to its closest regulated

[1] Unfortunately, access to the proprietary data on deregulated plants from EIA requires US citizenship.

neighbor and use the matched plant's coal price as its price. I am able to find a match for every deregulated plant within 100 miles. In the very few cases where no price data are available for a certain month, I use the average price of the months preceding and following that month instead. Table C.1 summarizes the matching outcomes. The final price for each plant is the quantity-weighted monthly receipt price.

Regulation Status	Coal Plant	Fuel Type	Matched Coal Plant
Deregulated	Big Brown	SUB	Gibbons Creek
		LIG	Pirkey
	Coleto Creek	SUB	J T Deely
	Limestone	SUB	Gibbons Creek
		LIG	Pirkey
	Martin Lake	SUB	Welsh
		LIG	Pirkey
	Monticello	SUB	Welsh
		LIG	Pirkey
	Oak Grove	LIG	Pirkey
	Sandow No 4	LIG	Pirkey
	Sandow No 5	LIG	Pirkey
	Twin Oaks Power One	LIG	Pirkey
	W A Parish	SUB	Fayette Power Project
Regulated	Gibbons Creek	SUB	
	Fayette Power Project	SUB	
	J K Spruce	SUB	
	J T Deely	SUB	
	Oklaunion	SUB	
	San Miguel	LIG	

Notes: This table shows the matching results for coal plants in ERCOT. As explained in the text, each deregulated plant is matched to a regulated plant that purchases the same type of coal.

Table C.1: Matching Outcomes for Coal Plants in ERCOT

C.2 Natural Gas Price

Daily natural gas spot prices are collected from SNL Financial. I use prices at the Agua Dulce, Katy, Waha, and Carthage hubs for units in the South, Houston, West, and North zones, respectively. Prices at the four hubs track each other very closely.

C.3 Variable Operation and Maintenance Costs (VOM)

Variable O&M costs include scheduled and forced outage maintenance, water supply costs, and environmental equipment maintenance. I use the standard VOM costs published by ERCOT (ERCOT, 2012). These costs differ by fuel and technology type. For coal, combined-cycle natural gas, natural gas combustion turbine and steam turbine , VOM costs are \$5.02, \$3.19, \$3.94 and \$7.08 per MWh (in 2009 dollars), respectively.

C.4 Emission Allowance Price

Power plants in ERCOT are subject to three programs: the Acid Rain Program (ARP), The Clean Air Interstate Rule (CAIR) annual SO_2 program and The Clean Air Interstate Rule (CAIR) annual NO_x program. The ARP, established under Title IV of the 1990 Clean Air Act (CAA) Amendments, requires major emission reductions of SO_2 and NO_x, the primary precursors of acid rain, from the power sector. It is a nationwide program affecting large fossil fuel-fired power plants across the country. CAIR was finalized in 2005, and took effect in 2009 for NO_x and 2010 for SO_2. The CAIR SO_2 and NO_x annual programs require further reductions for large electricity generating units in 28 eastern states including Texas. Not all generating units are affected by these three programs. To determine each generating units' coverage status, I use information provided by the EPA's Air Markets Program Data (AMPD) and cross check my data with The Code Of Federal Regulations Parts 72 and 96 (40 CFR Part 72 and Part 96).

All three programs are cap-and-trade programs designed to allow power plants to arrange for the cheapest possible reductions among covered sources to meet the overarching cap. For each ton of SO_2 emitted, ARP compliance requires the surrender of 1 ARP allowance, while CAIR compliance requires an additional ARP allowances of prompt vintage. For each ton of NO_x emitted, 1 NO_x annual allowance has to be deducted. Generally, these allowances are traded among companies and individuals through brokers. I acquire daily SO_2 and NO_x allowance price indexes from a leading over-the-counter energy brokerage firm based in Texas. I use the last trading price each day as the relevant price. For non-trading days, I approximate the price by taking the average of the prices from the two trading days preceding and following that day.

Compared to the fuel cost, the emission cost makes up a very small portion of the variable cost. For coal power plants, the emission cost on average counts for only 0.96%

of the marginal cost. For natural gas generators, the percentage is less than 0.3%.

C.5 Wholesale Electricity Price Data

From ERCOT, I also collect the real-time post-redesign electricity prices at four hubs: Houston, North, South and West. A hub's price is the simple average of the locational marginal prices (LMPs) of nodes within that hub.[2] When there is no congestion, the hub prices are the same throughout the system. However, if congestion does exist, LMPs differ from node to node, as do the hub prices. Therefore, I define an hour to be congested if the electricity prices at the four hubs are not the same. Congestion is quite common in my sample. Of all the hours in the post-redesign sample period, about 60% are congested.

[2]Locational marginal prices (LMPs) are prices at a given network node based on the cost of delivering the next MW of energy to that node. For example, if there is a need for 10 MW at a network node, the LMP would be determined by the cost of delivering the 11th MW.

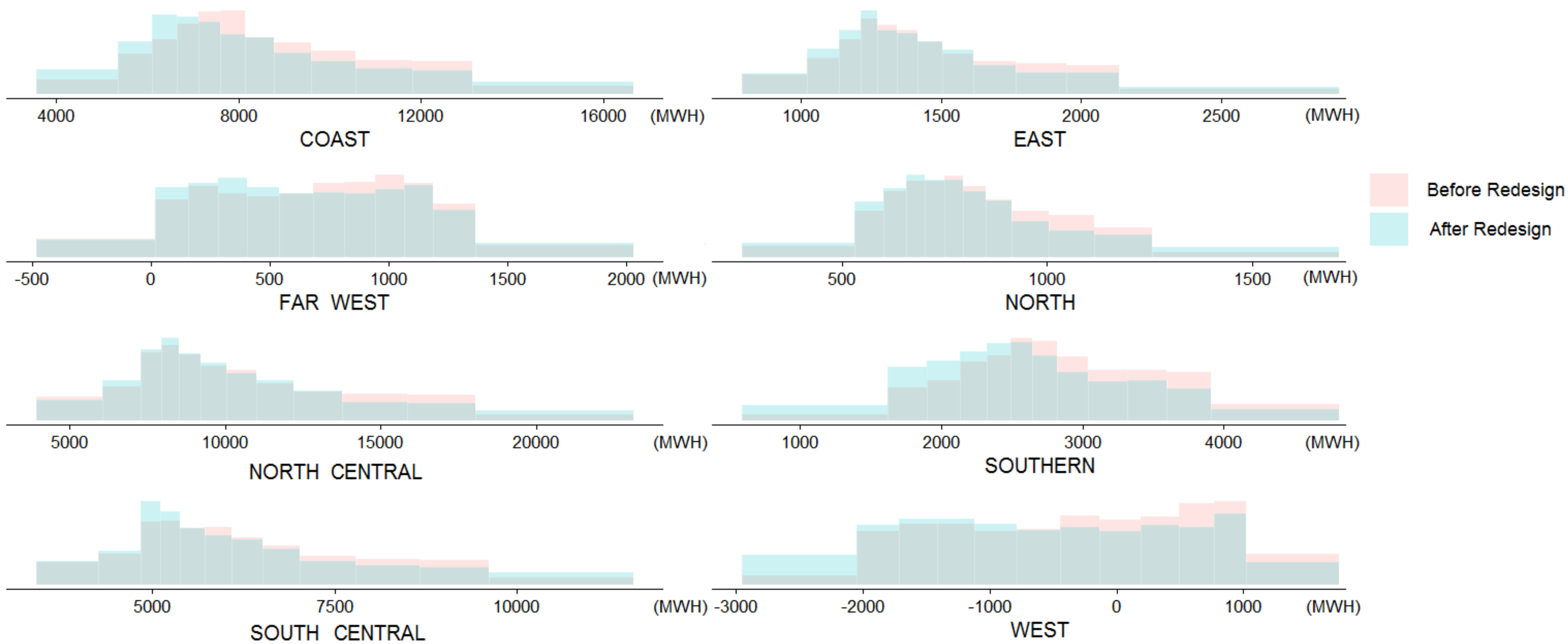
APPENDIX D

Additional Results and Robustness Checks

This section contains additional evidence that supports the validity of the empirical approach and the robustness of the findings. First, I show that although the market conditions pre- and post-redesign are not exactly the same, they are quite comparable: the span of demand at each zone overlaps; the changes in fuel prices are moderate; and the entry or exit of generators does not exert a significant impact on the market. Second, I provide further evidence showing that the observed changes pre- and post-redesign are not seen in any other year.

D.1 Comparison of Demand

Figure D.1 compares the distributions of demands at all eight weather zones pre- and post-redesign. To be consistent with the main specification, demand at each zone is divided into 12 equal-frequency bins based on the entire sample so that the number of observations falling into each bin is the same. Comparing the distributions before and after the redesign, we can see that they all have observations in each bin. The common support enables the estimation of the parameters for each bin both pre- and post-redesign.



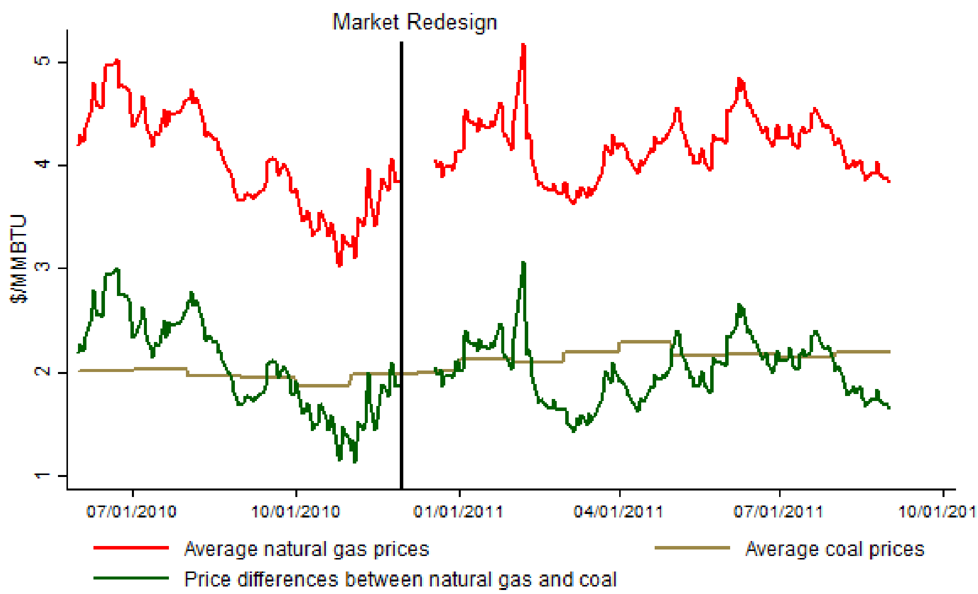
Notes: These figures show the histograms of hourly thermal demands at eight weather zones separately for pre-redesign and post-redesign periods. The pre-redesign period runs from June 1, 2010 to November 30, 2010. The post-redesign period runs from December 1, 2010 to August 31, 2011. Demand at each zone is divided into 12 equal-frequency bins based on the entire sample.

Figure D.1: Histogram of Hourly Thermal Demand By Weather Zone

D.2 Comparison of Fuel Prices

Changes in fuel prices are the only factor that may substantially affect generators' marginal costs. Other factors either do not change over time or constitute a very small portion of the total cost. In order to attribute the changes in generation to market redesign, it is essential to look at how fuel prices change during the sample period.

Figure D.2 plots the movement of coal prices and natural gas prices during the sample period.[1] Overall, the magnitudes of the price changes for both natural gas and coal are quite small. On average, coal and natural gas prices during the post-redesign period increase by 19.8 cents (10%) and 12.9 cents (3%) respectively, compared to the pre-redesign period. The ranges of the price differences are also similar pre- and post-redesign.



Notes: This figure shows the time series of average coal prices and natural gas prices as well as the price differences between coal and natural gas during the sample period which runs from June 1, 2010 to August 31, 2011 excluding December 1, 2010 to December 17, 2010 and February 2, 2011 to February 5, 2011. The vertical line indicates the time when the market redesign took place.

Figure D.2: Average Coal and Natural Gas Prices During the Sample Period

The comparability of fuel prices pre- and post-redesign provides reassuring evidence

[1] To be consist with the sample I use for estimation, I exclude dates between December 1, 2010 to December 17, 2010, for the lack of ERCOT data, and also dates between February 2, 2011 and February 5, 2011, for the unusual winter storm.

that my results are not driven by any price trend. Furthermore, I directly include a quadratic form of the price differences between natural gas and coal in the baseline model to capture any effect caused by relative changes in fuel prices. I find it is unlikely that the relative changes in fuel prices are the cause of the switch between coal and natural gas generation, because on average, the price for coal increases more than the price of natural gas during the post-redesign period. This would make coal generators less appealing, but instead I find coal displaces natural gas generation by significant amounts in the post-redesign period. Finally, as a robustness check, I focus on only natural gas generators and run similar regressions as equation (1.1) using demand for natural gas generation instead of thermal generation as the explanatory variable. Since within natural gas generators, the marginal cost order is basically determined by their heat rates and unaffected by the change of natural gas prices, any relative change in generation within them is not confounded by movements in fuel prices. The results from these regressions support the main findings: generation from cheaper resources, such as combined-cycle generators and combustion turbines, increases while generation from more costly resources, i.e. steam turbines, decreases.

D.3 Entry and Exit of Thermal Generators

In the estimation, I restrict the sample to all thermal generating units that were continually operating during the entire sample period. There are three thermal units that have either entered or exited the market in this time period. Two steam turbines, each with a capacity of 800 and 115 MW, exited the market prior to the market redesign. One combined-cycle natural gas plant with a capacity of 640 MW entered the market on March 16, 2011. These events pose the question of whether the observed changes in generation are caused by the entries or exits of these units. Although it is difficult to separate out their impact, I contend that this is not likely to be the case. The average hourly generation quantities from these three units while they were operating were only 69.9, 4.71 and 127.9 MWh, respectively. Their generation accounts for less than 0.1% of the total thermal generation. Given the magnitude of their average generation, I conclude that my results cannot be explained by their entry and exit.

D.4 Placebo Tests

Finally, I perform placebo tests to show that the magnitudes of the changes I find are indeed unusual and not seen in other years. For this exercise, I consider two hypothetical scenarios where a redesign occurred on December 1, 2009 as well as December 1, 2011 and repeat

Plant Name	Entry/Exit	Time	Technology	Capacity (MW)	Average Marginal Cost($)	Average Hourly Generation (MWh)
Tradinghouse	Exit	Sep 19, 2010	Natural gas: steam turbine	800	43.5	69.86
Permian Basin	Exit	Nov 21, 2010	Natural gas: steam turbine	115	49.4	4.71
Jack County	Entry	Mar 16, 2011	Natural gas: combined cycle	640	32.5	127.9

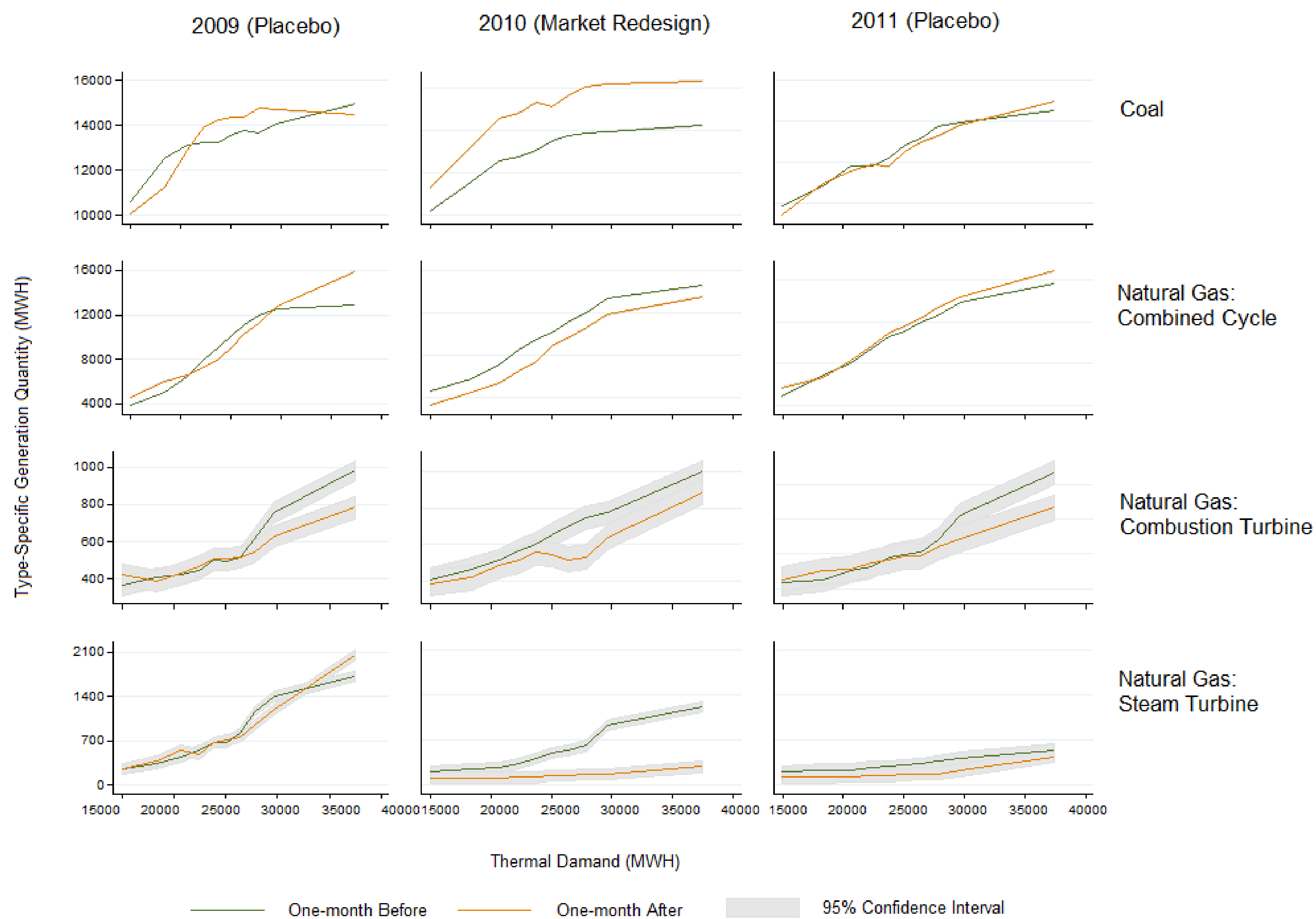
Notes: This table lists the thermal generators that have entered or exited the market during the sample period. The data come from EIA-860 forms. The entry/exit dates are also cross-checked with CEMS data and web sources.

Table D.1: List of Entering and Exiting Generators

the analyses. I focus on a very short time period – two months centered around the (real or pseudo) implementation date of the market redesign. Given the short time frame, I am confident that there are no significant changes in capacity, cost, or other aspects of the market. I run regressions similar to equation (1.1). However, given fewer observations, I simplify the analysis by using the entire thermal demand instead of the demand at each zone, and fitting a constant line within each bin. I also estimate the regressions at a more aggregate level by fuel and technology types. For type i at hour t, the estimation equation takes the following form:

$$\text{Gen}_{it} = \sum_{k=1}^{12} \beta_{ik} Bin_k(ThermalDemand_t) + \epsilon_{it}$$

where Bin_k is equal to one if the thermal demand falls into that bin and zero otherwise. Figure D.3 reports the results for the four categories: coal, combined-cycle natural gas, natural gas combustion turbines and steam turbines. From Figure D.3, we can see that the changes in 2010 are not seen in other years. For coal and combined-cycle natural gas, the “before” and “after” generation lines are intertwined and close to each other. Only in 2010 when there is a real redesign do we see significant gaps between these two lines. For combustion and steam turbines, the parameters are estimated less precisely. For steam turbines, there is evidence of a significant effect of market redesign when demand is over 20,000 MWh. Again, Figure D.3 shows that this is not seen in other years. For combustion turbines, the changes in 2010 are not very different from changes in 2009 and 2011. Overall, the 95% confidence intervals of the two generation lines overlap for 2010. However, this does not contradict the earlier findings that generation from combustion turbines increases, because the increase occurs only during high demand hours. Figure 1.7 shows that the effect starts to appear when the thermal demand exceeds 50,000 MWh. Given that the demand in November and December never reaches 50,000 MWh, it is unsurprising that the effect of market redesign on combustion turbines is not salient.



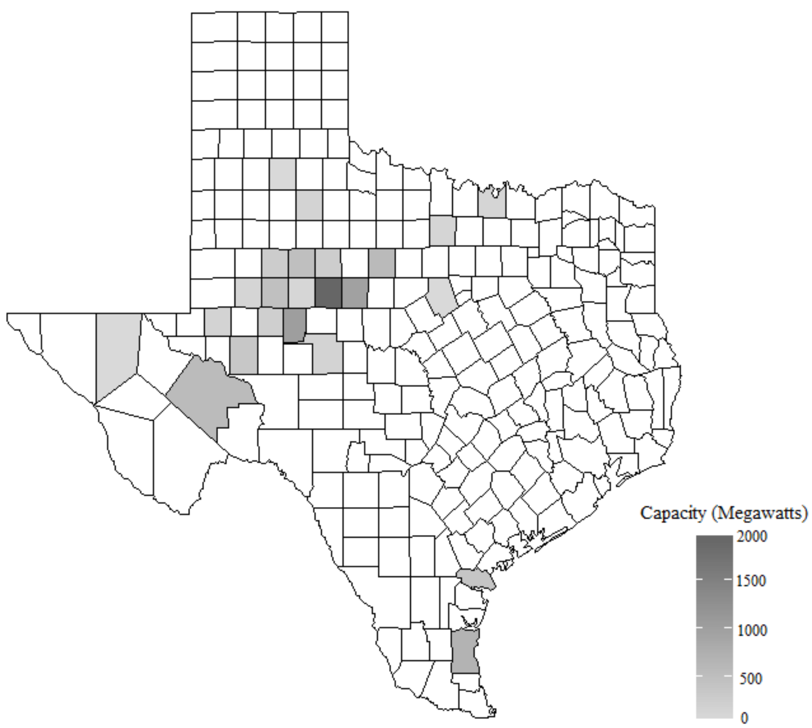
Notes: These figures report the results from the placebo tests as explained in the text.

Figure D.3: Results of Placebo Tests

APPENDIX E

The Effect of the Market Redesign on Wind Generation

Wind energy has a significant presence in the ERCOT region. As of 2010, installed wind capacity in ERCOT amounts to 9,363 MW, representing 9.45% of the overall generating capacity. As shown in Figure E.1, the majority of the wind farms are located in the west of Texas, with the rest in the southern portion of the state.



Notes: This figure is constructed by the author using data from the 2010 EIA-860 forms.

Figure E.1: Installed Wind Capacity in ERCOT: 2010

Wind power is determined by the availability of wind resources. Specifically, the level of electrical output a wind turbine can generate is proportional to its cross-sectional area as well as the cube of the wind speed. Wind is non-dispatchable in the sense that wind speed cannot be changed by will. However, cases do occur in which potential wind generation is not fully used. This happens largely because of the limited transmission capacity between western Texas where the most abundant wind resources are, and eastern Texas where most of the demand is. In some cases, wind generation has to be curtailed to avoid overloading the congested transmission lines.[1] Therefore, it is natural to ask if the market redesign results in fewer incidences of curtailment and better integration of wind resources.

Without data on the frequency of wind curtailments, I rely on a regression approach to examine the effect of market redesign on wind generation. The idea is that wind output is determined mostly by wind speed. If there is no effect, we should expect to see that the observed wind output curve stays more or less the same pre- and post-redesign. However, if redesign leads to better integration of wind resources, we should see a significant gap for wind outputs given the same wind speed and other market conditions before and after the redesign. During the sample period, 350 MW of additional wind capacity was added. To rule out this effect, I restrict my sample to a subset of wind farms that were already in operation as of June 1, 2010. I conduct the analysis at the weather zone level. For zone i at hour t, I estimate the following regression:

$$GEN_{it} = \theta_i After_t + \sum_{k=1}^{3} \alpha_{ik} WSP_{it}^{(k)} + \sum_{k=1}^{8} \beta_{ik} Demand_{kt} + \delta_h + \epsilon_{it}$$

where GEN_{it} is the aggregated wind generation quantity in zone i at hour t. $After$ is a dummy that indicates the post-redesign period. WSP is the average wind speed cubed.[2] I also include demands in all eight weather zones as well as hourly fixed effects. Newey-West standard errors are calculated using 24-hour lags.

Table E.1 shows the results of the coefficients for $After$ at the five weather zones which have non-zero installed wind capacity. Although there appear to be some increases in generation when only wind speed is included in the model, this effect goes away as more

[1] See Sioshansi and Hurlbut (2010) for an extensive discussion of the ERCOT market protocols with respect to wind generation.

[2] Wind speed data are collected from National Centers for Environmental Information(NCEI)'s Integrated Surface Database. One station is selected from each county where wind farms exist and data are available. The average wind speed for each zone is calculated by taking the simple average of the stations within that zone.

Weather Zone	Model (1)	Model (2)	Model (3)
Far West	53.01**	31.30	-19.98
	(24.03)	(23.35)	(26.15)
North	21.54***	18.34***	2.94
	(5.17)	(4.76)	(5.57)
North Central	26.72***	12.42	-26.19**
	(9.95)	(9.18)	(12.10)
Southern	7.68	8.42	-13.56
	(8.98)	(9.05)	(12.03)
West	23.97	-44.42	-129.56***
	(36.28)	(30.45)	(36.04)
Cube of WSP	Y	Y	Y
Hour	N	Y	Y
Demand	N	N	Y

Notes: Newey-West standard errors are reported in the parentheses.
*** Significant at the 1% confidence level
** Significant at the 5% confidence level

Table E.1: Effect of the Market Redesign on Wind Generation

controls are added. In the full model, there is no evidence of a significant increase in wind generation after the redesign. If anything, the results seem to suggest that wind generation is lower in some of the regions post-redesign.

BIBLIOGRAPHY

[1] Acharya, Viral, and Alberto Bisin. "Counterparty Risk Externality: Centralized versus Over-the-counter Markets." *Journal of Economic Theory* 149, (2014): 153-182.

[2] Alberini, A. and D. Austin (2002): "Accidents Waiting to Happen: Liability Policy and Toxic Pollution Releases." *The Review of Economics and Statistics* 84, no. 4(2002): 729-741.

[3] Almond, Douglas, Yuyu Chen, Michael Greenstone and Hongbin Li. "Winter Heating or Clean Air? Unintended Impacts of China's Huai River Policy." *American Economic Review: Papers & Proceedings* 99, (2009): 184-190.

[4] Andrews, Steven Q. "Inconsistencies in Air Quality Metrics: "Blue Sky" Days and PM_{10} Concentrations in Beijing." *Environmental Research Letters* 3, (2008): 1-14.

[5] Auffhammer, Maximilian, and Ryan Kellogg."Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality." *American Economic Review* 101, (2011): 2687-2722.

[6] Ben-Shahar, Omri and K. D. Logue. "Outsourcing Regulation: How Insurance Reduces Moral Hazard." *Michigan Law Review* 111, (2012): 197–248.

[7] Boomhower, Judson. "Drilling like There's No Tomorrow: Bankruptcy, Insurance, and Environmental Risk." *Haas Working Paper* no. 254 (2016).

[8] Borenstein, Severin, James Bushnell, and Steven Stoft. "The Competitive Effects of Transmission Capacity in A Deregulated Electricity Industry." *RAND Journal of Economics* 31, no. 2 (2000): 294-325.

[9] Borenstein, Severin, James B. Bushnell, and Frank A. Wolak. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review* 92, no. 5 (2002): 1376-1405.

[10] Born, P. and D. Chen. "Coping with Environmental Risks: Development of a Market for Environmental Liability Insurance in China." *Working Paper* Available at http://orientation.uchicagolawandecon.org/readings/colloquia/Chen%20-%20Coping%20with%20Environmental%20Risk.pdf (2013).

[11] Cardell, Judith B., Carrie Cullen Hitt, and William W. Hogan. "Market Power and Strategic Interaction in Electricity Networks." *Resource and Energy Economics* 19, no. 1 (1997): 109-137.

[12] Chen, Yihsu and Alexander Whalley. "Green Infrastructure: The Effects of Urban Rail Transit on Air Quality." *American Economic Journal: Economic Policy* 4, no. 1 (2012): 58-97.

[13] Chen, Yuyu, Ginger Zhe Jin, Naresh Kumar and Guang Shi."Gaming in Air Pollution Data? Lessons from China." *The B.E. Journal of Economic Analysis & Policy* 13, no. 3 (2012): 1-43.

[14] China Environment Yearbook Editing Committee. "China Environment Yearbook." *China Environment Yearbook Press, Beijing* (2007-2015).

[15] Chu, Yin, Scott Holladay, and Jacob LaRiviere. "Pass-through from Fossil Fuel Market Prices to Procurement Costs of the US Power Producers." Available at http://web.utk.edu/~jhollad3/PassThrough.pdf (2015).

[16] Cicala, Steve. "Imperfect Markets versus Imperfect Regulation in U.S. Electricity Generation." *NBER Working Paper* no. 23053 (2017).

[17] Cohen, A. and R. Dehejia. "The Effect of Automobile Insurance and Accident Liability Laws on Traffic Fatalities." *Journal of Law and Economics* XLVII, (2004).

[18] Congjun, L. and A. BinChik. "Developing Environmental Pollution Liability Insurance: the Countermeasures – A Case Study of Baoding, Hebei Province, China." Available at http://www.ccirm.org/conference/2012/uploadfiles/A/II-A/2-Liu_Congjun-Developing%20Environmental%20Pollution%20Liability%20Insurance%20the%20Countermeasures.pdf (2012).

[19] CRA International and Resero Consulting. "Update on the ERCOT Nodal Market Cost-Benefit Analysis." Available at http://www.puc.texas.gov/industry/electric/reports/31600/PUCT_CBA_Report_Final.pdf (2008).

[20] Cullen, Joseph A., and Erin T. Mansur. "Inferring Carbon Abatement Costs in Electricity Markets: A revealed Preference Approach Using the Shale Revolution." *NBER Working Paper* no. w20795 (2014).

[21] Davis, Lucas W. "The Effect of Driving Restrictions on Air Quality in Mexico City." *Journal of Political Economy* 116, (2008): 38-81.

[22] Davis, Lucas W., and Catherine Hausman. "Market Impacts of a Nuclear Power Plant Closure." *American Economic Journal: Applied Economics* 8, no. 2 (2016): 92-122.

[23] Dewatripont, Mathias, and Eric Maskin. "Credit and Efficiency in Centralized and Decentralized Economies." *The Review of Economic Studies* 62, no. 4 (1995): 541-555.

[24] Electric Reliability Council of Texas (ERCOT). "Market Benefits Observed The First 6 Months." Available at http://www.ercot.com/content/meetings/board/keydocs/2011/0621/Item_06_-_Market_Benefits_Report.pdf (2011a).

[25] Electric Reliability Council of Texas (ERCOT). "Financial Summary." Available at http://www.ercot.com/content/meetings/board/keydocs/2011/0322/Item_05a_-_Financial_Summary_Report.pdf (2011b).

[26] Electric Reliability Council of Texas (ERCOT). Available at http://www.ercot.com/content/meetings/wms/keydocs/2010/0324/13_white_paper_vcwg_proposed_alt_to_vc_100317.doc (2012).

[27] Energy Information Administration (EIA). "Electric Power Annual." Available at https://www.eia.gov/electricity/annual/archive/2011/ (2011).

[28] Environmental Protection Agency (EPA). "Regulatory Impact Analysis Control of Air Pollution from New Motor Vehicles: Tier 2 Motor Vehicle Emissions Standards and Gasoline Sulfur Control Requirements." *Washington, D.C.: U.S. Environmental Protection Agency* (1999).

[29] Environmental Protection Agency (EPA). "Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866 (May 2013, Revised August 2016)." Available at https://www.epa.gov/sites/production/files/2016-12/documents/sc_co2_tsd_august_2016.pdf (2016).

[30] Farber, S. “Regulatory Schemes and Self-protective Environmental Risk Control: a Comparison of Insurance, Liability, and Deposit/ Refund Systems.” *Ecological Economics* 3, (1991): 231–245.

[31] Feng, Y., A. Mol, Y. Lu, G. He, and C. VanKoppen. “Environmental Pollution Liability Insurance in China In Need of Strong Government Backing.” *AMBIO* 43, no. 5 (2014): 687-702.

[32] Ghanem, Dalia, and Junjie Zhang.“Effortless Perfection: Do Chinese Cities Manipulate “Blue Skies”?” *Journal of Environmental Economics and Management* 68, no. 2 (2013): 203-225.

[33] Glode, Vincent, and Christian Opp. “Can Decentralized Markets Be More Efficient?” Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2697281` (2016).

[34] Hahn, Jinyong, Petra Todd and Wilbert Van der Klaauw. “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” *Econometrica* 69, (2001): 201-209.

[35] Hogan, William W. “Coordination for Competition in an Electricity Market.” *John F. Kennedy School of Government, Harvard University, Cambridge, MA* (1995).

[36] Imbens G, Lemieux T. “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Econometrics* 142, no. 2 (2008): 615-635.

[37] Jaramillo, Paulina, and Nicholas Z. Muller. “The Air Pollution Damage from Energy Production in the U.S: 2002 - 2011.” *Energy Policy* 90, (2016): 202-211.

[38] Jha, Akshaya. “Regulatory Induced Risk Aversion: Coal Procurement at U.S. Power Plants.” Available at `http://web.stanford.edu/~akshayaj/reg_risk_aversion.pdf` (2014).

[39] Joskow, Paul L. “Contract Duration and Relationship-specific Investments: Empirical Evidence from Coal Markets.” *American Economic Review* 77, no. 1 (1987): 168-185.

[40] Joskow, Paul L. “Deregulation and Regulatory Reform in the US Electric Power Sector.” *Deregulation of Network Industries: The Next Steps* (2000).

[41] Joskow, Paul L., and Jean Tirole. “Transmission Rights and Market Power on Electric Power Networks.” *Rand Journal of Economics* 31, no. 3 (2000): 450-487.

[42] Joskow, P. L., and E. Kahn. "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000." *Energy Journal* 23, no. 4 (2002): 1-35.

[43] Katzman, M. T. "Pollution Liability Insurance and Catastrophic Environmental Risk." *Journal of Risk and Insurance* 55, no. 1 (1988): 55-75.

[44] Kolstad, C. D., T. S. Ulen, and G. V. Johnson. "Ex Post Liability for Harm vs. Ex Ante Safety Regulation: Substitutes or Complements?" *American Economic Review* 80, (1990): 888-901.

[45] Lai, Hak-Kan, Hilda Tsang and Chit-Ming Wong. "Meta-Analysis of Adverse Health Effects Due to Air Pollution in Chinese Populations" *BMC Public Health* 13, (2013): 360.

[46] Liu, Huan et al. "Analysis of the Impacts of Fuel Sulfur on Vehicle Emissions in China." *Fuel* 87, (2008): 3147-3154.

[47] Manufacturers of Emission Controls Association. "The Impact of Gasoline Fuel Sulfur on Catalytic Emission Control Systems." Available at www.meca.org/galleries/files/sulfur.pdf (1998).

[48] Mansur, Erin T. "Measuring Welfare in Restructured Electricity Markets." *The Review of Economics and Statistics* 90, no. 2 (2008): 369-386.

[49] Mansur, Erin T and Matthew W. White. "Market Organization and Efficiency in Electricity Markets." Available at http://www.dartmouth.edu/~mansur/papers/mansur_white_pjmaep.pdf (2012).

[50] Marion, Justin and Erich Muehlegger. "Measuring Illegal Activity and the Effects of Regulatory Innovation: Tax Evasion and the Dyeing of Untaxed Diesel." *Journal of Political Economy* 116, no. 4 (2008): 633-666.

[51] Ministry of Environmental Protection (MEP) and China Insurance Regulatory Commission (CIRC). "The Guidelines on Environmental Pollution Liability Insurance." *no. 189* (2007).

[52] Ministry of Environmental Protection (MEP) and China Insurance Regulatory Commission (CIRC). "The Guiding Opinions on Pilot Scheme for Compulsory Environmental Pollution Liability Insurance." *no. 10* (2013).

[53] Petit, A., Jeffrey, J., Palmer, F., and Steinbrink, R. "European Programme on Emissions, Fuels and Engine Technologies (EPEFE)-Emissions from Gasoline Sulphur Study." *SAE Technical Paper* no. 961071 (1996).

[54] Potomac Economics. "2011 State of the Market Report for the ERCOT Wholesale Electricity Markets." Available at https://www.puc.texas.gov/industry/electric/reports/ERCOT_annual_reports/2011annualreport.pdf (2012).

[55] Reguant, Mar. "Complementary Bidding Mechanisms and Startup Costs in Electricity Markets." *The Review of Economic Studies* 81, no. 4 (2014): 1708-1742.

[56] Rogerson, W. P. "Repeated Moral Hazard." *Econometrica* 53, (1985): 69-76.

[57] Rubinstein, A. and M. E. Yaari. "Repeated Insurance Contracts and Moral Hazard." *Journal of Economic Theory* 30, (1983): 74-97.

[58] Ryan, Nicholas. "The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market." *NBER Working Paper* no. 23106 (2017).

[59] Sioshansi, Ramteen, and David Hurlbut. "Market Protocols in ERCOT and Their Effect on Wind Generation." *Energy Policy* 38, no. 7 (2010): 3192-3197.

[60] Sisler, James F., and William C. Malm. "The Relative Importance of Soluble Aerosols to Spatial and Seasonal Trends of Impaired Visibility in the United States." *Atmosphere Environment* 28, (1994): 851-862.

[61] Trans-Energy Research Associates, China Petrochemical Consulting Corporation, and Lawrence Berkeley National Laboratory. "Improving Transport Fuel Quality in China: Implications for the Refining Sector." Available at https://eetd.lbl.gov/sites/all/files/publications/lbl-50681-improving-tranport-fuelaugust2002.pdf (2002).

[62] Walsh, Michael P. "A Handbook on Low Sulfur Fuels in China." Available at http://cleanairinitiative.org/portal/sites/default/files/documents/a_handbook_on_low_sulfur_fuels_in_china-final_0.pdf (2013).

[63] Wang, Hua and Jie He. "The Value of Statistical Life: A Contingent Investigation in China." *World Bank: Policy Research Working Paper* no. 5421 (2010).

[64] Wilson, Robert. “Architecture of Power Markets.” *Econometrica* 70, no. 4 (2002): 1299-1340.

[65] Wolak, Frank A., and Robert H. Patrick. “The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market.” *NBER Working Paper* no. w8248 (2001).

[66] Wolak, Frank A. “Measuring the Competitiveness Benefits of a Transmission Investment Policy: The Case of the Alberta Electricity Market.” *Energy Policy* 85, (2015): 426-444.

[67] Wolfram, Catherine D. “Measuring Duopoly Power in the British Electricity Spot Market.” *American Economic Review* 89, no. 4 (1999): 805-826.

[68] Wolinsky, Asher. “Information Revelation in a Market with Pairwise Meetings.” *Econometrica* 58, no. 1 (1990): 1-23.

[69] Yin, H., A. Pfaff, and H. Kunreuther. “Can Environmental Insurance Succeed Where Other Strategies Fail? The Case of Underground Storage Tanks.” *Risk Analysis* 31, no. 1 (2011): 12-24.

[70] Zhang, Qingfeng and Robert Crooks. “Toward an Environmentally Sustainable Future: Country Environmental Analysis of the People’s Republic of China.” *Mandaluyong City, Philippines: Asian Development Bank* (2012).

[71] Zhang, Yu et al. “Systematic Review of Chinese Studies of Short-term Exposure to Air Pollution and Daily Mortality.” *Environment International* 54, (2013): 100-111.

[72] Zweifel, P. and J. R. Tyran. “Environmental Impairment Liability as an Instrument of Environmental Policy.” *Ecological Economics* 11, (1993): 43–56.